Regularized Learning under Reproducing Kernel Banach Spaces: Similarity and Feature Representations

Jun Zhang

University of Michigan, Ann Arbor

In collaboration with

Haizhang Zhang, Yuesheng Xu, Matt Jones

May 20, 2017, SCNU, China

Fruitful Collaboration with H Zhang and Y Xu

- Zhang, H., Xu, Y., and Zhang, J. (2009) Reproducing kernel Banach spaces for machine learning. *Journal of Machine Learning Research.* 10: 2741-2775.
- Zhang, H. and Zhang, J. (2010). Generalized semi-inner products with application to regularized learning. *Journal of Mathematical Analysis and Application.* 372: 181-196.
- Zhang, H. and Zhang, J. (2011). Frames, Riesz bases, and sampling expansions in Banach spaces via semi-inner products. *Applied and Computational Harmonic Analysis.* 31:1-25.
- Zhang, H. and Zhang, J. (2012). Regularized learning in Banach space as an optimization problem: Representer theorems. *Journal of Global Optimization.* 54: 235-250
- Zhang, H. and Zhang, J. (2013). Vector-valued Reproducing Kernel Banach Spaces with applications to multi-task learning, *Journal of Complexity.* 29: 195-215.
- Zhang, H. and Zhang, J. (2015). Learning with Reproducing Kernel Banach Spaces. *Proceedings of the 10th ISAAC Congress, Macau.* In *New Trends in Analysis and Interdisciplinary Applications.* Birkhauser.
- Zhang, H. and Zhang, J. (2017). Learning with reproducing kernel Banach spaces. New Trends in Analysis and Interdisciplinary Applications. Springer.

Learning: Regularization and Optimization

Learning-from-data in an underconstrained (ill-posed) problem, so needs to be regularized. Find the minimizer that provides better goodness-of-fit but lower complexity:

$$\inf_{f \in \mathcal{B}} \ \mathcal{L}_{\mathbf{z}}(f) + \lambda \phi(\|f\|_{\mathcal{B}}),$$

where

- $\mathcal{B}$: a Banach space of candidate functions;
- $\mathcal{L}_{\mathbf{z}}$: "loss" function measuring how well the sample data $\mathbf{z} = \{(x_i, y_i)_{i \in \mathcal{I}}\}$ is fitted by the input-output relation $\{x_i, f(x_i)\}_{i \in \mathcal{I}}\}$ by a candidate function f;
- λ: Lagrange multiplier regularizing the balance of two competing forces for optimal generalization;
- ϕ: a non-decreasing function "modulating" the "capacity" of the function space $\mathcal{B}$ where the minimizer may lie.

The Case with Reproducing Kernel Hilbert Space

- A well-understood case of regularized learning is the Reproducing Kernel Hilbert Space (RKHS):

$$\min_{f \in \mathcal{H}_K} \mathcal{L}_{\mathbf{z}}(f) + \lambda\phi(\|f\|_{\mathcal{H}_K})$$

where $\mathcal{H}_K$ is the RKHS with the associated kernel K.

- Here K, called the "reproducing kernel" of $\mathcal{H}$, is a symmetric function on $X \times X$ such that $K(x,\cdot) \in \mathcal{H}, K(\cdot,x) \in \mathcal{H}$ for all $x \in X$ and

$$f(x) = \langle f(\cdot), K(x,\cdot)\rangle_{\mathcal{H}}, \ x \in X$$

where $\langle , \rangle$ denotes the inner product.

The Case with Reproducing Kernel Hilbert Space

- A well-understood case of regularized learning is the Reproducing Kernel Hilbert Space (RKHS):

$$\min_{f \in \mathcal{H}_K} \mathcal{L}_{\mathbf{z}}(f) + \lambda \phi(\|f\|_{\mathcal{H}_K})$$

  where $\mathcal{H}_K$ is the RKHS with the associated kernel K.

- Here K, called the "reproducing kernel" of $\mathcal{H}$, is a symmetric function on $X \times X$ such that $K(x, \cdot) \in \mathcal{H}, K(\cdot, x) \in \mathcal{H}$ for all $x \in X$ and

$$f(x) = \langle f(\cdot), K(x, \cdot) \rangle_{\mathcal{H}}, \;\; x \in X$$

  where $\langle , \rangle$ denotes the inner product.

Properties of Kernels in RKHS

Reproducing Property: $K : X \times X \to \mathbb{C}$ satisfies

$$K(x, y) = \langle K(x, \cdot), K(y, \cdot) \rangle_{\mathcal{H}_K}.$$

Representer Theorem: nature of the solution

$$f(\cdot) = \sum_{i=1}^{n} c_i K(x_i, \cdot), \text{ for some } c_i \in \mathbb{C}, 1 \leq i \leq n.$$

Characterization Theorem: nature of the kernel

A symmetric function K is a reproducing kernel iff

1. There exists a feature mapping $\Phi : X \to \mathcal{W}$ such that
$$K(x, y) = \langle \Phi(x), \Phi(y) \rangle_{\mathcal{W}}, \quad x, y \in X; \qquad \textbf{Or}$$
2. For any sequence $\{x_i\}_{i \in \mathcal{I}}$, kernel matrix $\mathbb{K}_{ij} = K(x_i, x_j)$ is positive semi-definite, i.e., for any $c_i, i \in \mathcal{I}$
$$\sum_{i,j} K(x_i, x_j) c_i \bar{c}_j \geq 0.$$

Properties of Kernels in RKHS

Reproducing Property: $K : X \times X \to \mathbb{C}$ satisfies

$K(x, y) = \langle K(x, \cdot), K(y, \cdot) \rangle_{\mathcal{H}_K}$.

Representer Theorem: nature of the solution

$$f(\cdot) = \sum_{i=1}^{n} c_i K(x_i, \cdot), \text{ for some } c_i \in \mathbb{C}, 1 \le i \le n.$$

Characterization Theorem: nature of the kernel

A symmetric function K is a reproducing kernel iff

1. There exists a feature mapping $\Phi : X \to \mathcal{W}$ such that
$$K(x, y) = \langle \Phi(x), \Phi(y) \rangle_{\mathcal{W}}, \quad x, y \in X; \qquad \textbf{Or}$$
2. For any sequence $\{x_i\}_{i \in \mathcal{I}}$, kernel matrix $\mathbb{K}_{ij} = K(x_i, x_j)$ is positive semi-definite, i.e., for any $c_i, i \in \mathcal{I}$
$$\sum_{i,j} K(x_i, x_j) c_i \bar{c}_j \ge 0.$$

Properties of Kernels in RKHS

Reproducing Property: $K : X \times X \to \mathbb{C}$ satisfies

$$K(x, y) = \langle K(x, \cdot), K(y, \cdot)\rangle_{\mathcal{H}_K}.$$

Representer Theorem: nature of the solution

$$f(\cdot) = \sum_{i=1}^{n} c_i K(x_i, \cdot), \text{ for some } c_i \in \mathbb{C}, 1 \le i \le n.$$

Characterization Theorem: nature of the kernel

A symmetric function K is a reproducing kernel iff

1. There exists a feature mapping $\Phi : X \to \mathcal{W}$ such that
$$K(x, y) = \langle \Phi(x), \Phi(y)\rangle_{\mathcal{W}}, \quad x, y \in X; \qquad \textbf{Or}$$
2. For any sequence $\{x_i\}_{i \in \mathcal{I}}$, kernel matrix $\mathbb{K}_{ij} = K(x_i, x_j)$ is positive semi-definite, i.e., for any $c_i, i \in \mathcal{I}$
$$\sum_{i,j} K(x_i, x_j) c_i \bar{c}_j \ge 0.$$

Properties of Kernels in RKHS

Reproducing Property: $K : X \times X \to \mathbb{C}$ satisfies

$$K(x, y) = \langle K(x, \cdot), K(y, \cdot)\rangle_{\mathcal{H}_K}.$$

Representer Theorem: nature of the solution

$$f(\cdot) = \sum_{i=1}^{n} c_i K(x_i, \cdot), \text{ for some } c_i \in \mathbb{C}, 1 \leq i \leq n.$$

Characterization Theorem: nature of the kernel

A symmetric function K is a reproducing kernel iff

1. There exists a feature mapping $\Phi : X \to \mathcal{W}$ such that
$$K(x, y) = \langle \Phi(x), \Phi(y)\rangle_{\mathcal{W}}, \quad x, y \in X; \qquad \textbf{Or}$$
2. For any sequence $\{x_i\}_{i \in \mathcal{I}}$, kernel matrix $\mathtt{K}_{ij} = K(x_i, x_j)$ is positive semi-definite, i.e., for any $c_i, i \in \mathcal{I}$
$$\sum_{i,j} K(x_i, x_j) c_i \bar{c}_j \geq 0.$$

RKHS as Unifying Model of Human Categorization

1. (Exemplar model) people judge the similarity of a test item to all remembered exemplars of each category (Nosofsky, 1986);
2. (Prototype model) people judge the similarity of a test item to a prototype of each category, representing the average feature values of all category members (Smith and Minda, 1998);
3. (Decision-bound model) people learn boundaries in stimulus space that separate categories (Maddox and Ashby, 1993);
4. (Perceptron model) people learn associations between individual stimulus features and category labels with an error-driven mechanism (Gluck and Bower, 1988);
5. (ALCOVE model) associative learning is coupled with learning to selectively attend to relevant dimensions as individual exemplars are presented (Kruschke, 1992);
6. (Cluster model) categories as mixture distributions, with each component of this mixture given by a Gaussian centered on some hypothetical stimulus (Anderson, 1991).

RKHS as Unifying Model of Human Categorization

1. (Exemplar model) people judge the similarity of a test item to all remembered exemplars of each category (Nosofsky, 1986);
2. (Prototype model) people judge the similarity of a test item to a prototype of each category, representing the average feature values of all category members (Smith and Minda, 1998);
3. (Decision-bound model) people learn boundaries in stimulus space that separate categories (Maddox and Ashby, 1993);
4. (Perceptron model) people learn associations between individual stimulus features and category labels with an error-driven mechanism (Gluck and Bower, 1988);
5. (ALCOVE model) associative learning is coupled with learning to selectively attend to relevant dimensions as individual exemplars are presented (Kruschke, 1992);
6. (Cluster model) categories as mixture distributions, with each component of this mixture given by a Gaussian centered on some hypothetical stimulus (Anderson, 1991).

RKHS as Unifying Model of Human Categorization

1. (Exemplar model) people judge the similarity of a test item to all remembered exemplars of each category (Nosofsky, 1986);
2. (Prototype model) people judge the similarity of a test item to a prototype of each category, representing the average feature values of all category members (Smith and Minda, 1998);
3. (Decision-bound model) people learn boundaries in stimulus space that separate categories (Maddox and Ashby, 1993);
4. (Perceptron model) people learn associations between individual stimulus features and category labels with an error-driven mechanism (Gluck and Bower, 1988);
5. (ALCOVE model) associative learning is coupled with learning to selectively attend to relevant dimensions as individual exemplars are presented (Kruschke, 1992);
6. (Cluster model) categories as mixture distributions, with each component of this mixture given by a Gaussian centered on some hypothetical stimulus (Anderson, 1991).

RKHS as Unifying Model of Human Categorization

1. (Exemplar model) people judge the similarity of a test item to all remembered exemplars of each category (Nosofsky, 1986);
2. (Prototype model) people judge the similarity of a test item to a prototype of each category, representing the average feature values of all category members (Smith and Minda, 1998);
3. (Decision-bound model) people learn boundaries in stimulus space that separate categories (Maddox and Ashby, 1993);
4. (Perceptron model) people learn associations between individual stimulus features and category labels with an error-driven mechanism (Gluck and Bower, 1988);
5. (ALCOVE model) associative learning is coupled with learning to selectively attend to relevant dimensions as individual exemplars are presented (Kruschke, 1992);
6. (Cluster model) categories as mixture distributions, with each component of this mixture given by a Gaussian centered on some hypothetical stimulus (Anderson, 1991).

RKHS as Unifying Model of Human Categorization

1. (Exemplar model) people judge the similarity of a test item to all remembered exemplars of each category (Nosofsky, 1986);
2. (Prototype model) people judge the similarity of a test item to a prototype of each category, representing the average feature values of all category members (Smith and Minda, 1998);
3. (Decision-bound model) people learn boundaries in stimulus space that separate categories (Maddox and Ashby, 1993);
4. (Perceptron model) people learn associations between individual stimulus features and category labels with an error-driven mechanism (Gluck and Bower, 1988);
5. (ALCOVE model) associative learning is coupled with learning to selectively attend to relevant dimensions as individual exemplars are presented (Kruschke, 1992);
6. (Cluster model) categories as mixture distributions, with each component of this mixture given by a Gaussian centered on some hypothetical stimulus (Anderson, 1991).

RKHS as Unifying Model of Human Categorization

1. (Exemplar model) people judge the similarity of a test item to all remembered exemplars of each category (Nosofsky, 1986);
2. (Prototype model) people judge the similarity of a test item to a prototype of each category, representing the average feature values of all category members (Smith and Minda, 1998);
3. (Decision-bound model) people learn boundaries in stimulus space that separate categories (Maddox and Ashby, 1993);
4. (Perceptron model) people learn associations between individual stimulus features and category labels with an error-driven mechanism (Gluck and Bower, 1988);
5. (ALCOVE model) associative learning is coupled with learning to selectively attend to relevant dimensions as individual exemplars are presented (Kruschke, 1992);
6. (Cluster model) categories as mixture distributions, with each component of this mixture given by a Gaussian centered on some hypothetical stimulus (Anderson, 1991).

RKHS as Unifying Model of Human Categorization

1. (Exemplar model) people judge the similarity of a test item to all remembered exemplars of each category (Nosofsky, 1986);
2. (Prototype model) people judge the similarity of a test item to a prototype of each category, representing the average feature values of all category members (Smith and Minda, 1998);
3. (Decision-bound model) people learn boundaries in stimulus space that separate categories (Maddox and Ashby, 1993);
4. (Perceptron model) people learn associations between individual stimulus features and category labels with an error-driven mechanism (Gluck and Bower, 1988);
5. (ALCOVE model) associative learning is coupled with learning to selectively attend to relevant dimensions as individual exemplars are presented (Kruschke, 1992);
6. (Cluster model) categories as mixture distributions, with each component of this mixture given by a Gaussian centered on some hypothetical stimulus (Anderson, 1991).

RKHS as Unifying Model of Human Categorization

1. (Exemplar model) people judge the similarity of a test item to all remembered exemplars of each category (Nosofsky, 1986);
2. (Prototype model) people judge the similarity of a test item to a prototype of each category, representing the average feature values of all category members (Smith and Minda, 1998);
3. (Decision-bound model) people learn boundaries in stimulus space that separate categories (Maddox and Ashby, 1993);
4. (Perceptron model) people learn associations between individual stimulus features and category labels with an error-driven mechanism (Gluck and Bower, 1988);
5. (ALCOVE model) associative learning is coupled with learning to selectively attend to relevant dimensions as individual exemplars are presented (Kruschke, 1992);
6. (Cluster model) categories as mixture distributions, with each component of this mixture given by a Gaussian centered on some hypothetical stimulus (Anderson, 1991).

Can Kernels be Asymmetric?

1. Symmetry of kernels is tied to inner-product of a Hilbert space. So a major challenge of generalizing kernel methods is to deal with the lack of "inner-product" in a Banach space $\mathcal{B}$.
2. Strategy: invoke duality mapping between $\mathcal{B} \leftrightarrow \mathcal{B}^*$, where $\mathcal{B}^*$ is the space of continuous linear functionals on $\mathcal{B}$ with the norm given by

$$||g^*||_* = \sup_{f:||f||=1} g^*(f) = \sup_f \frac{(f, g^*)}{||f||}$$

   where $(\,,\,) : \mathcal{B} \times \mathcal{B}^* \to \mathbb{C}$ denotes a bilinear form called "duality pairing".
3. Solution: "Semi-inner-product" operator on $\mathcal{B}$, a natural generalization of "inner-product" on $\mathcal{H}$.

Can Kernels be Asymmetric?

1. Symmetry of kernels is tied to inner-product of a Hilbert space. So a major challenge of generalizing kernel methods is to deal with the lack of "inner-product" in a Banach space $\mathcal{B}$.
2. Strategy: invoke duality mapping between $\mathcal{B} \leftrightarrow \mathcal{B}^*$, where $\mathcal{B}^*$ is the space of continuous linear functionals on $\mathcal{B}$ with the norm given by

$$||g^*||_* = \sup_{f:||f||=1} g^*(f) = \sup_f \frac{(f, g^*)}{||f||}$$

   where $(,) : \mathcal{B} \times \mathcal{B}^* \to \mathbb{C}$ denotes a bilinear form called "duality pairing".
3. Solution: "Semi-inner-product" operator on $\mathcal{B}$, a natural generalization of "inner-product" on $\mathcal{H}$.

Can Kernels be Asymmetric?

1. Symmetry of kernels is tied to inner-product of a Hilbert space. So a major challenge of generalizing kernel methods is to deal with the lack of "inner-product" in a Banach space $\mathcal{B}$.
2. Strategy: invoke duality mapping between $\mathcal{B} \leftrightarrow \mathcal{B}^*$, where $\mathcal{B}^*$ is the space of continuous linear functionals on $\mathcal{B}$ with the norm given by

$$||g^*||_* = \sup_{f:||f||=1} g^*(f) = \sup_f \frac{(f, g^*)}{||f||}$$

where $(,) : \mathcal{B} \times \mathcal{B}^* \to \mathbb{C}$ denotes a bilinear form called "duality pairing".

3. Solution: "Semi-inner-product" operator on $\mathcal{B}$, a natural generalization of "inner-product" on $\mathcal{H}$.

Can Kernels be Asymmetric?

1. Symmetry of kernels is tied to inner-product of a Hilbert space. So a major challenge of generalizing kernel methods is to deal with the lack of "inner-product" in a Banach space $\mathcal{B}$.
2. Strategy: invoke duality mapping between $\mathcal{B} \leftrightarrow \mathcal{B}^*$, where $\mathcal{B}^*$ is the space of continuous linear functionals on $\mathcal{B}$ with the norm given by

$$||g^*||_* = \sup_{f:||f||=1} g^*(f) = \sup_f \frac{(f, g^*)}{||f||}$$

   where $(,) : \mathcal{B} \times \mathcal{B}^* \to \mathbb{C}$ denotes a bilinear form called "duality pairing".
3. Solution: "Semi-inner-product" operator on $\mathcal{B}$, a natural generalization of "inner-product" on $\mathcal{H}$.

Semi-Inner-Product Operator

- (Lumer, 1961) A *semi-inner-product* $[\cdot,\cdot]$ on $\mathcal{B}\times\mathcal{B}$ satisfies, for all $f,g,h\in\mathcal{B}$ and $\alpha\in\mathbb{C}$
  1. $[f+g,h]=[f,h]+[g,h]$, $[\alpha f,g]=\alpha[f,g]$;
  2. $[f,f]>0$ for $f\neq 0$;
  3. (Cauchy-Schwartz) $|[f,g]|^2\leq[f,f][g,g]$.
- From (2) and (3), a norm is induced: $\|f\|_{\mathcal{B}}=[f,f]^{1/2}$.
- Giles (1967) shows that $[f,\alpha g]=\bar{\alpha}[f,g]$.
- If $\overline{[g,f]}=[f,g]$, then $[f,g+h]=[f,g]+[f,h]$.
- Nath (1971) showed that (3) can be generalized to define the s.i.p of order p (satisfying Hölder inequality):

$$|[f,g]|\leq([f,f])^{1/p}([g,g])^{(p-1)/p}$$

In this case $[f,f]^{1/p}=\|f\|_{\mathcal{B}}$.

Semi-Inner-Product Operator

- (Lumer, 1961) A *semi-inner-product* $[\cdot,\cdot]$ on $\mathcal{B}\times\mathcal{B}$ satisfies, for all $f,g,h\in\mathcal{B}$ and $\alpha\in\mathbb{C}$
  1. $[f+g,h]=[f,h]+[g,h]$, $[\alpha f,g]=\alpha[f,g]$;
  2. $[f,f]>0$ for $f\neq 0$;
  3. (Cauchy-Schwartz) $|[f,g]|^2\leq[f,f][g,g]$.
- From (2) and (3), a norm is induced: $\|f\|_{\mathcal{B}}=[f,f]^{1/2}$.
- Giles (1967) shows that $[f,\alpha g]=\bar{\alpha}[f,g]$.
- If $\overline{[g,f]}=[f,g]$, then $[f,g+h]=[f,g]+[f,h]$.
- Nath (1971) showed that (3) can be generalized to define the s.i.p of order p (satisfying Hölder inequality):

$$|[f,g]|\leq([f,f])^{1/p}([g,g])^{(p-1)/p}$$

In this case $[f,f]^{1/p}=\|f\|_{\mathcal{B}}$.

Semi-Inner-Product Operator

- (Lumer, 1961) A *semi-inner-product* $[\cdot,\cdot]$ on $\mathcal{B}\times\mathcal{B}$ satisfies, for all $f, g, h \in \mathcal{B}$ and $\alpha \in \mathbb{C}$
  1. $[f+g,h] = [f,h]+[g,h]$, $[\alpha f, g] = \alpha[f,g]$;
  2. $[f,f] > 0$ for $f \neq 0$;
  3. (Cauchy-Schwartz) $|[f,g]|^2 \leq [f,f][g,g]$.
- From (2) and (3), a norm is induced: $\|f\|_{\mathcal{B}} = [f,f]^{1/2}$.
- Giles (1967) shows that $[f,\alpha g] = \bar{\alpha}[f,g]$.
- If $\overline{[g,f]} = [f,g]$, then $[f,g+h] = [f,g]+[f,h]$.
- Nath (1971) showed that (3) can be generalized to define the s.i.p of order p (satisfying Hölder inequality):

$$|[f,g]| \leq ([f,f])^{1/p}([g,g])^{(p-1)/p}$$

In this case $[f,f]^{1/p} = \|f\|_{\mathcal{B}}$.

Semi-Inner-Product Operator

- (Lumer, 1961) A *semi-inner-product* $[\cdot,\cdot]$ on $\mathcal{B}\times\mathcal{B}$ satisfies, for all $f,g,h\in\mathcal{B}$ and $\alpha\in\mathbb{C}$
  1. $[f+g,h]=[f,h]+[g,h]$, $[\alpha f,g]=\alpha[f,g]$;
  2. $[f,f]>0$ for $f\neq 0$;
  3. (Cauchy-Schwartz) $|[f,g]|^2\leq[f,f][g,g]$.
- From (2) and (3), a norm is induced: $\|f\|_{\mathcal{B}}=[f,f]^{1/2}$.
- Giles (1967) shows that $[f,\alpha g]=\bar{\alpha}[f,g]$.
- If $\overline{[g,f]}=[f,g]$, then $[f,g+h]=[f,g]+[f,h]$.
- Nath (1971) showed that (3) can be generalized to define the s.i.p of order p (satisfying Hölder inequality):

$$|[f,g]|\leq([f,f])^{1/p}([g,g])^{(p-1)/p}$$

In this case $[f,f]^{1/p}=\|f\|_{\mathcal{B}}$.

Semi-Inner-Product Operator

- (Lumer, 1961) A *semi-inner-product* $[\cdot,\cdot]$ on $\mathcal{B}\times\mathcal{B}$ satisfies, for all $f,g,h\in\mathcal{B}$ and $\alpha\in\mathbb{C}$
  1. $[f+g,h]=[f,h]+[g,h]$, $[\alpha f,g]=\alpha[f,g]$;
  2. $[f,f]>0$ for $f\neq 0$;
  3. (Cauchy-Schwartz) $|[f,g]|^2\leq[f,f][g,g]$.
- From (2) and (3), a norm is induced: $\|f\|_{\mathcal{B}}=[f,f]^{1/2}$.
- Giles (1967) shows that $[f,\alpha g]=\bar{\alpha}[f,g]$.
- If $\overline{[g,f]}=[f,g]$, then $[f,g+h]=[f,g]+[f,h]$.
- Nath (1971) showed that (3) can be generalized to define the s.i.p of order p (satisfying Hölder inequality):

$$|[f,g]|\leq([f,f])^{1/p}([g,g])^{(p-1)/p}$$

In this case $[f,f]^{1/p}=\|f\|_{\mathcal{B}}$.

Semi-Inner-Product Operator (cont)

Semi-inner-product allows us to express

1. continuous linear functional: $g^*(f) = [f, g]$.
2. pseudo-orthogonality: $[f, g] = 0$ iff $f \perp g$ (in the sense of James) $||g + tf|| > ||g||$ for all $t \neq 0$.
3. asymmetric projection and hence angles between vectors: $\cos\theta_{f,g} = |[f, g]|/(||f|| \cdot ||g||^{p-1})$.

Existence and uniqueness of semi-inner-product

If $\mathcal{B}$ is uniformly convex and uniformly Fréchet differentiable:

$$\mathrm{Re}([f, g]) = \frac{1}{p} \lim_{t \to 0} \frac{||g + tf||^p - ||g||^p}{t}.$$

Example. For $\mathcal{B} = L^p(X, \mu)$, its s.i.p. is

$$[f, g] = \frac{\int f \bar{g} |g|^{p-2} d\mu}{||g||_p^{p-2}}.$$

Semi-Inner-Product Operator (cont)

Semi-inner-product allows us to express

1. continuous linear functional: $g^*(f) = [f, g]$.
2. pseudo-orthogonality: $[f, g] = 0$ iff $f \perp g$ (in the sense of James) $||g + tf|| > ||g||$ for all $t \neq 0$.
3. asymmetric projection and hence angles between vectors: $\cos\theta_{f,g} = |[f, g]|/(||f|| \cdot ||g||^{p-1})$.

Existence and uniqueness of semi-inner-product

If $\mathcal{B}$ is uniformly convex and uniformly Fréchet differentiable:

$$\mathrm{Re}([f, g]) = \frac{1}{p} \lim_{t \to 0} \frac{||g + tf||^p - ||g||^p}{t}.$$

Example. For $\mathcal{B} = L^p(X, \mu)$, its s.i.p. is

$$[f, g] = \frac{\int f \bar{g} |g|^{p-2} d\mu}{||g||_p^{p-2}}.$$

Semi-Inner-Product Operator (cont)

Semi-inner-product allows us to express

1. continuous linear functional: $g^*(f) = [f, g]$.
2. pseudo-orthogonality: $[f, g] = 0$ iff $f \perp g$ (in the sense of James) $||g + tf|| > ||g||$ for all $t \neq 0$.
3. asymmetric projection and hence angles between vectors: $\cos\theta_{f,g} = |[f, g]|/(||f|| \cdot ||g||^{p-1})$.

Existence and uniqueness of semi-inner-product

If $\mathcal{B}$ is uniformly convex and uniformly Fréchet differentiable:

$$\mathrm{Re}([f, g]) = \frac{1}{p} \lim_{t \to 0} \frac{||g + tf||^p - ||g||^p}{t}.$$

Example. For $\mathcal{B} = L^p(X, \mu)$, its s.i.p. is

$$[f, g] = \frac{\int f \bar{g} |g|^{p-2} d\mu}{||g||_p^{p-2}}.$$

Semi-Inner-Product Operator (cont)

Semi-inner-product allows us to express

1. continuous linear functional: $g^*(f) = [f, g]$.
2. pseudo-orthogonality: $[f, g] = 0$ iff $f \perp g$ (in the sense of James) $||g + tf|| > ||g||$ for all $t \neq 0$.
3. asymmetric projection and hence angles between vectors: $\cos\theta_{f,g} = |[f, g]|/(||f|| \cdot ||g||^{p-1})$.

Existence and uniqueness of semi-inner-product

If $\mathcal{B}$ is uniformly convex and uniformly Fréchet differentiable:

$$\mathrm{Re}([f, g]) = \frac{1}{p} \lim_{t\to 0} \frac{||g + tf||^p - ||g||^p}{t}.$$

Example. For $\mathcal{B} = L^p(X, \mu)$, its s.i.p. is

$$[f, g] = \frac{\int f\bar{g}|g|^{p-2} d\mu}{||g||_p^{p-2}}.$$

Semi-Inner-Product Operator (cont)

Semi-inner-product allows us to express

1. continuous linear functional: $g^*(f) = [f, g]$.
2. pseudo-orthogonality: $[f, g] = 0$ iff $f \perp g$
   (in the sense of James) $||g + tf|| > ||g||$ for all $t \neq 0$.
3. asymmetric projection and hence angles between vectors:
   $\cos\theta_{f,g} = |[f, g]|/(||f|| \cdot ||g||^{p-1})$.

Existence and uniqueness of semi-inner-product

If $\mathcal{B}$ is uniformly convex and uniformly Fréchet differentiable:

$$\mathrm{Re}([f, g]) = \frac{1}{p} \lim_{t \to 0} \frac{||g + tf||^p - ||g||^p}{t}.$$

Example. For $\mathcal{B} = L^p(X, \mu)$, its s.i.p. is

$$[f, g] = \frac{\int f \bar{g} |g|^{p-2} d\mu}{||g||_p^{p-2}}.$$

Generalized Semi-Inner-Products

We can generalize s.i.p further.
Let $\varphi : \mathbb{R}_+ \to \mathbb{R}_+$ be strictly increasing, $\varphi(0) = 0$, $\lim_{t\to\infty} \varphi(t) = \infty$. Denote $\psi(t) := t/\varphi(t)$.

Definition 1. Generalized Semi-Inner-Product

A g.s.i.p. $[\cdot,\cdot]_\varphi \to \mathbb{C}$ on $\mathcal{B} \times \mathcal{B}$ satisfies, for all $f, g, h \in \mathcal{B}$ and $\alpha \in \mathbb{C}$

1. $[f + g, h]_\varphi = [f, h]_\varphi + [g, h]_\varphi$, $[\alpha f, g]_\varphi = \alpha[f, g]_\varphi$;
2. $[f, f]_\varphi > 0$ for $f \neq 0$;
3. $|[f, g]_\varphi| \leq \varphi([f, f])\psi([g, g])$.

Generalized Semi-Inner-Products

We can generalize s.i.p further.
Let $\varphi : \mathbb{R}_+ \to \mathbb{R}_+$ be strictly increasing, $\varphi(0) = 0$, $\lim_{t\to\infty} \varphi(t) = \infty$. Denote $\psi(t) := t/\varphi(t)$.

Definition 1. Generalized Semi-Inner-Product

A g.s.i.p. $[\cdot,\cdot]_\varphi \to \mathbb{C}$ on $\mathcal{B} \times \mathcal{B}$ satisfies, for all $f, g, h \in \mathcal{B}$ and $\alpha \in \mathbb{C}$

1. $[f+g,h]_\varphi = [f,h]_\varphi + [g,h]_\varphi$, $[\alpha f, g]_\varphi = \alpha[f,g]_\varphi$;
2. $[f,f]_\varphi > 0$ for $f \neq 0$;
3. $|[f,g]_\varphi| \leq \varphi([f,f])\psi([g,g])$.

G.s.i.p. and Generalized Duality Mapping

Condition (3) replaces Cauchy-Schwartz (and the more general Hölder) inequality. When $\varphi(t) = t^{1/p}$ and $\psi(t) = t^{1/q}$, where $p, q > 1$, $1/p + 1/q = 1$, ours reduces to s.i.p of type p (Nath, 1971) which, for $p = q = 2$, reduces to that of Lumer (1961).

Proposition 1 (Zhang and Zhang, 2010)

1. G. s.i.p induces a norm on $\mathcal{B}$ by $\|f\| = \varphi([f, f]_\varphi)$.
2. For each fixed $f \in \mathcal{B}$, $[\cdot, f]_\varphi$ defines a continuous linear functional on $\mathcal{B}$, and therefore is an element of $\mathcal{B}^*$.
3. Further, $[\cdot, f]_\varphi$ can be identified with the dual element f^* defined by the generalized duality mapping:

$$f^*(f) = ||f^*||_* ||f||, \quad \text{and} \quad ||f^*||_* = \gamma(||f||).$$

where $\gamma(t) = \varphi^{-1}(t)/t$.

Duality Mapping in Reflexive Banach Spaces

Let Γ be a Young function on $[0,+\infty)$, i.e., Γ is an increasing, strictly convex function with $\Gamma(0)=0$, and let Γ^* be its convex conjugate, i.e.,
$\Gamma^*(s)=\sup_{t>0}(ts-\Gamma(t))=s(\Gamma')^{-1}(s)-\Gamma((\Gamma')^{-1}(s))$.

Lemma

The pair of functions $\tilde{\Gamma},\tilde{\Gamma}^*$ given by

$$\tilde{\Gamma}(f):=\Gamma(\|f\|_{\mathcal{B}}) \text{ and } \tilde{\Gamma}^*(g^*):=\Gamma^*(\|g^*\|_{\mathcal{B}^*}),\quad f\in\mathcal{B}, g^*\in\mathcal{B}^*$$

are strictly convex functions on $\mathcal{B}$ and $\mathcal{B}^*$, respectively, and that are conjugate to each other. They satisfy

$$\Gamma(\|f\|_{\mathcal{B}})+\Gamma^*(\|g^*\|_{\mathcal{B}^*})\geq\|f\|_{\mathcal{B}}\,\|g^*\|_{\mathcal{B}^*}\geq|(f,g^*)|.$$

- The first inequality is due to the strictly convexity of Γ; equality holds when the norms satisfy $\|g^*\|_{\mathcal{B}^*} = \Gamma'(\|f\|_{\mathcal{B}}) = \gamma(||f||_{\mathcal{B}})$.
- The second inequality is due to the definition of dual norm; equality holds when g^* lies in a certain direction (denoted f^*) in $\mathcal{B}^*$ dual to the direction of $f \in \mathcal{B}$.
- The two equalities hold simultaneously iff $g^* = f^*$ where

$$f^* \equiv [\cdot, f]_\varphi \equiv J_\gamma(f)$$

where $J_\gamma : \mathcal{B} \to \mathcal{B}^*, f \mapsto J_\gamma(f)$ denotes a generalized duality mapping with "gauge" γ:

$$\|J_\gamma(f)\|_{\mathcal{B}^*} = \gamma(\|f\|_{\mathcal{B}}),\ J_\gamma(f)(f) = f^*(f) = \varphi^{-1}(||f||_{\mathcal{B}}).$$

Note that J_γ as a function from $\mathcal{B}$ to $\mathcal{B}^*$ is generally non-linear $J_\gamma(f_1 + f_2) \neq J_\gamma(f_1) + J_\gamma(f_2)$.

Duality Mapping in Reflexive Banach Space (Cont)

- The first inequality is due to the strictly convexity of Γ; equality holds when the norms satisfy $\|g^*\|_{\mathcal{B}^*} = \Gamma'(\|f\|_{\mathcal{B}}) = \gamma(||f||_{\mathcal{B}})$.
- The second inequality is due to the definition of dual norm; equality holds when g^* lies in a certain direction (denoted f^*) in $\mathcal{B}^*$ dual to the direction of $f \in \mathcal{B}$.
- The two equalities hold simultaneously iff $g^* = f^*$ where

$$f^* \equiv [\cdot, f]_\varphi \equiv J_\gamma(f)$$

where $J_\gamma : \mathcal{B} \to \mathcal{B}^*, f \mapsto J_\gamma(f)$ denotes a generalized duality mapping with "gauge" γ:

$$\|J_\gamma(f)\|_{\mathcal{B}^*} = \gamma(\|f\|_{\mathcal{B}}), \; J_\gamma(f)(f) = f^*(f) = \varphi^{-1}(||f||_{\mathcal{B}}).$$

Note that J_γ as a function from $\mathcal{B}$ to $\mathcal{B}^*$ is generally non-linear $J_\gamma(f_1 + f_2) \neq J_\gamma(f_1) + J_\gamma(f_2)$.

Duality Mapping in Reflexive Banach Space (Cont)

- The first inequality is due to the strictly convexity of Γ; equality holds when the norms satisfy $\|g^*\|_{\mathcal{B}^*} = \Gamma'(\|f\|_{\mathcal{B}}) = \gamma(||f||_{\mathcal{B}})$.
- The second inequality is due to the definition of dual norm; equality holds when g^* lies in a certain direction (denoted f^*) in $\mathcal{B}^*$ dual to the direction of $f \in \mathcal{B}$.
- The two equalities hold simultaneously iff $g^* = f^*$ where

$$f^* \equiv [\cdot, f]_\varphi \equiv J_\gamma(f)$$

  where $J_\gamma : \mathcal{B} \to \mathcal{B}^*, f \mapsto J_\gamma(f)$ denotes a generalized duality mapping with "gauge" γ:

$$\|J_\gamma(f)\|_{\mathcal{B}^*} = \gamma(\|f\|_{\mathcal{B}}),\ J_\gamma(f)(f) = f^*(f) = \varphi^{-1}(||f||_{\mathcal{B}}).$$

Note that J_γ as a function from $\mathcal{B}$ to $\mathcal{B}^*$ is generally non-linear $J_\gamma(f_1 + f_2) \neq J_\gamma(f_1) + J_\gamma(f_2)$.

Duality Mapping in Reflexive Banach Space (Cont)

- Written explicitly,

$$J_\gamma(f) = \frac{\gamma(\|f\|_{\mathcal{B}})}{\|f\|_{\mathcal{B}}} f_0^* = \frac{\varphi^{-1}(\|f\|_{\mathcal{B}})}{\|f\|_{\mathcal{B}}^2} f_0^*$$

  where the conventional dual f_0^* satisfies $\|f_0^*\|_{\mathcal{B}^*} = \|f\|_{\mathcal{B}}$, corresponding to $\Gamma(t) = 1/2t^2$. Here

$$\varphi^{-1}(t) := t\Gamma'(t) = t\gamma(t), \quad t \in [0, +\infty).$$

- The special case of Nath's s.i.p. of order p, with $\varphi(t) = t^{\frac{1}{p}}, \psi(t) = t^{\frac{1}{q}}$, corresponds to

$$\varphi^{-1}(t) = t^p, \quad \gamma(t) = t^{p-1}, \quad \Gamma(t) = \frac{1}{p} t^p$$

.

Reproducing Kernel Banach Spaces (RKBS)

We *define* an RKBS as a reflexive Banach space $\mathcal{B}$ of functions on X such that the dual space $\mathcal{B}^*$ is also a Banach space of functions on X and point evaluations are continuous linear functionals on both $\mathcal{B}$ and $\mathcal{B}^*$.

Theorem 2 (Zhang, Xu, and Zhang, 2009)

Let $\mathcal{B}$ be an RKBS on X. Then there exists a function K on $X \times X$ (called the "reproducing kernel" of $\mathcal{B}$) such that $K(x, \cdot) \in \mathcal{B}$, $K(\cdot, x) \in \mathcal{B}^*$ for all $x \in X$ and

$$f(x) = (f, K(\cdot, x)), \quad f^*(x) = (K(x, \cdot), f^*), \; f \in \mathcal{B}, \; f^* \in \mathcal{B}^*.$$

Here $(\cdot, \cdot)$ is a bilinear form ("dual pairing") on $\mathcal{B} \times \mathcal{B}^*$:

$$(f, g^*) := g^*(f), \quad f \in \mathcal{B}, \; g^* \in \mathcal{B}^*.$$

Reproducing Kernel Banach Spaces (RKBS)

We *define* an RKBS as a reflexive Banach space $\mathcal{B}$ of functions on X such that the dual space $\mathcal{B}^*$ is also a Banach space of functions on X and point evaluations are continuous linear functionals on both $\mathcal{B}$ and $\mathcal{B}^*$.

Theorem 2 (Zhang, Xu, and Zhang, 2009)

Let $\mathcal{B}$ be an RKBS on X. Then there exists a function K on $X \times X$ (called the "reproducing kernel" of $\mathcal{B}$) such that $K(x, \cdot) \in \mathcal{B}$, $K(\cdot, x) \in \mathcal{B}^*$ for all $x \in X$ and

$$f(x) = (f, K(\cdot, x)), \quad f^*(x) = (K(x, \cdot), f^*), \; f \in \mathcal{B}, \; f^* \in \mathcal{B}^*.$$

Here $(\cdot, \cdot)$ is a bilinear form ("dual pairing") on $\mathcal{B} \times \mathcal{B}^*$:

$$(f, g^*) := g^*(f), \quad f \in \mathcal{B}, \; g^* \in \mathcal{B}^*.$$

Reproducing Kernel Banach Spaces (Cont)

In general $K(x,y) \neq K(y,x)$, but they satisfy

$$K(x,y) = (K(x,\cdot), K(\cdot,y)).$$

Characterization Theorem (Zhang, Xu, and Zhang, 2009)

A bivariate function $K : X \times X \to \mathbb{C}$ is a reproducing kernel in a RKBS iff there exist mappings Φ from X to some reflexive Banach space $\mathcal{W}$ and $\Phi^* : X \to \mathcal{W}^*$ such that

$$K(x,y) = (\Phi(x), \Phi^*(y))_{\mathcal{W}}, \quad x, y \in X.$$

where $(,)_{\mathcal{W}}$ denotes the pairing of $\mathcal{W}$ with $\mathcal{W}^*$.

Reproducing Kernel Banach Spaces (Cont)

In general $K(x, y) \neq K(y, x)$, but they satisfy

$$K(x, y) = (K(x, \cdot), K(\cdot, y)).$$

Characterization Theorem (Zhang, Xu, and Zhang, 2009)

A bivariate function $K : X \times X \to \mathbb{C}$ is a reproducing kernel in a RKBS iff there exist mappings Φ from X to some reflexive Banach space $\mathcal{W}$ and $\Phi^* : X \to \mathcal{W}^*$ such that

$$K(x, y) = (\Phi(x), \Phi^*(y))_{\mathcal{W}}, \quad x, y \in X.$$

where $(,)_{\mathcal{W}}$ denotes the pairing of $\mathcal{W}$ with $\mathcal{W}^*$.

Reproducing Kernel Banach Spaces (Cont)

In general $K(x, y) \neq K(y, x)$, but they satisfy

$$K(x, y) = (K(x, \cdot), K(\cdot, y)).$$

Characterization Theorem (Zhang, Xu, and Zhang, 2009)

A bivariate function $K : X \times X \to \mathbb{C}$ is a reproducing kernel in a RKBS iff there exist mappings Φ from X to some reflexive Banach space $\mathcal{W}$ and $\Phi^* : X \to \mathcal{W}^*$ such that

$$K(x, y) = (\Phi(x), \Phi^*(y))_{\mathcal{W}}, \quad x, y \in X.$$

where $(,)_{\mathcal{W}}$ denotes the pairing of $\mathcal{W}$ with $\mathcal{W}^*$.

Semi-Inner-Product (S.i.p) Kernel

When $\mathcal{B}$ also admits an semi-inner-product (i.e., uniformly convex and uniformly Fréchet differentiable), we construct $G(x, \cdot) \equiv (K(\cdot, x))^* \in \mathcal{B}$ and write the reproducing kernel as:

S.i.p. Reproducing Kernel (Zhang, Xu, and Zhang, 2009)

Let $\mathcal{B}$ be a RKBS on X with given a s.i.p., and K its reproducing kernel. Then there exists a unique function $G : X \times X \to \mathbb{C}$ such that $\{G(x, \cdot) : x \in X\} \subseteq \mathcal{B}$ and

$$f(x) = [f, G(x, \cdot)], \quad \text{for all } f \in \mathcal{B}, x \in X.$$

The s.i.p. kernel G satisfies $G(x, y) = [G(x, \cdot), G(y, \cdot)]$. It is characterized by the existence of a feature map $\Phi : X \to W$ such that $G(x, y) = [\Phi(x), \Phi(y)]_W$.

Representer Theorems for RKBS

Let $\mathcal{B}$ be an RKBS with reproducing kernel K. Consider the regularized learning

$$\inf_{f \in \mathcal{B}} \mathcal{L}_{\mathbf{z}}(f) + \lambda \phi(\|f\|_{\mathcal{B}}). \tag{1}$$

Representer Theorem (Zhang and Zhang, 2011)

Suppose that (1) has at least one minimizer. If ϕ is *strictly increasing*, then every minimizer f_0 of (1) must satisfy, for some complex constants c_j, $1 \leq j \leq n$,

$$f_0^*(\cdot) = \sum_{j=1}^{n} c_j K(\cdot, x_j).$$

If ϕ is *nondecreasing*, then there exists at least one minimizer of (1) that has the above form.

Representer Theorems for RKBS

Let $\mathcal{B}$ be an RKBS with reproducing kernel K. Consider the regularized learning

$$\inf_{f\in\mathcal{B}} \mathcal{L}_{\mathbf{z}}(f) + \lambda\phi(\|f\|_{\mathcal{B}}). \tag{1}$$

Representer Theorem (Zhang and Zhang, 2011)

Suppose that (1) has at least one minimizer. If ϕ is *strictly increasing*, then every minimizer f_0 of (1) must satisfy, for some complex constants c_j, $1 \le j \le n$,

$$f_0^*(\cdot) = \sum_{j=1}^{n} c_j K(\cdot, x_j).$$

If ϕ is *nondecreasing*, then there exists at least one minimizer of (1) that has the above form.

Dual Representations

Duality Between Object and Feature

1. An object has *dual* representations:
   - in the input space (where processing is described by exemplars and their similarity);
   - in the feature space (where processing is described by feature-outcome associations).
2. Similarity and feature represenation as two-sides of the categorization "coin":
   - similarity-based generalization (in Hilbert space)
   - sparsity-based dimension reduction and compressed representation (in l_1 space)

Dual Representations

Duality Between Object and Feature

1. An object has *dual* representations:
   - in the input space (where processing is described by exemplars and their similarity);
   - in the feature space (where processing is described by feature-outcome associations).
2. Similarity and feature represenation as two-sides of the categorization "coin":
   - similarity-based generalization (in Hilbert space)
   - sparsity-based dimension reduction and compressed representation (in l_1 space)

Symbolic Representation of Object vs Feature

Modeling Object-Feature as Cross-Table

In a cross-table $V \times E$, a binary relation R between a set V ("objects") and another set E ("features") is represented by the entry of the table, such that an entry is 1 whenever object v possesses feature e, and entry is 0 otherwise. So "feature" serves to classify objects.

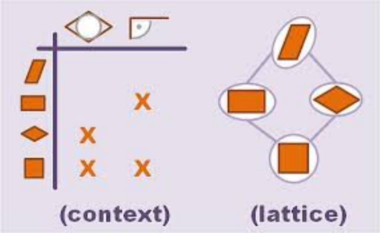


Figure: Cross-table

Relating Cross-Table to Hypergraph

In a hypergraph (V, E), each element $v \in V$ is a "vertex", and each member e of E (e being a subset of V) is a "hyperedge" (i.e., an "edge" connecting multiple vertices).

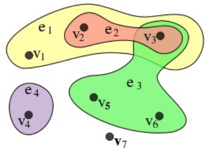


Figure: Hypergraph

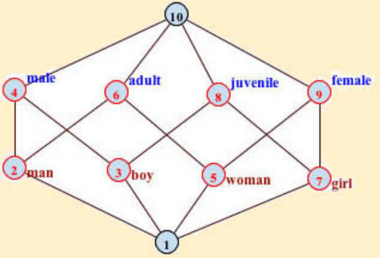
Formal Concept Analysis and Concept Lattice

Starting from a cross-table, there is a formal procedure to derive a so-called "concept lattice", where each node corresponds to:

1. the (maximal) collection of objects sharing a given list of features;
2. the (maximal) list of features shared by a given collection of objects.

Fca	female	juvenile	adult	male
girl	x	x	0	0
woman	x	0	x	0
boy	0	x	0	x
man	0	0	x	x

LIST OF FORMAL CONCEPTS
C1 {},{female,juvenile,adult,male}
C2 {man},{adult,male}
C3 {boy},{juvenile,male}
C4 {boy,man},{male}
C5 {woman},{female,adult}
C6 {woman,man},{adult}
C7 {girl},{female,juvenile}
C8 {girl,boy},{juvenile}
C9 {girl,woman},{female}
C10 {girl,woman,boy,man},{}



Closure as an Operator

For any cross-table (V, E), the closure operation is established through the Galois connection – they form a "Concept Lattice", which is a semi-modular lattice.

Axiomatization of Closure Operation

A mapping $\mathbf{Cl} : \mathcal{P}(\mathbf{V}) \to \mathcal{P}(\mathbf{V})$ is a closure operator if **Cl** satisfies the following axioms:
$(\mathbf{C1})$ $A \subseteq \mathbf{Cl}(\mathbf{A})$;
$(\mathbf{C2})$ $A \subseteq B \to \mathbf{Cl}(\mathbf{A}) \subseteq \mathbf{Cl}(\mathbf{B})$;
$(\mathbf{C3})$ $\mathbf{Cl}(\mathbf{Cl}(\mathbf{A})) = \mathbf{Cl}(\mathbf{A})$.
Those A that satisfies $A = \mathbf{Cl}(\mathbf{A})$ is called "closed" set.

Note that: subspaces of Hilbert space forms a modular lattice (projective geometry). Ongoing work (with Dr. Yinbin Lei) investigates the algebraic structures of "Knowlede Space" and "Learning Space".

Boundary Point

Given a cross-table (V, E), we define two types of points with respect to any non-empty subset $A \subseteq V$ (which does not have to be a member of E):

Boundary Point

A point $v \in V$ (not necessarily in A) is called A's *boundary point* if for every $e \in E(v)$,

$$e \cap A \neq \emptyset, \quad e \cap A^c \neq \emptyset \; (\leftrightarrow e \setminus A \neq \emptyset).$$

The set of all boundary points of A is denoted $\mathbf{Fr(A)}$.

From its definition,

$$\mathbf{Fr(A)} = \mathbf{Fr(A^c)}.$$

Boundary Point

Given a cross-table (V, E), we define two types of points with respect to any non-empty subset $A \subseteq V$ (which does not have to be a member of E):

Boundary Point

A point $v \in V$ (not necessarily in A) is called A's *boundary point* if for every $e \in E(v)$,

$$e \cap A \neq \emptyset, \quad e \cap A^c \neq \emptyset \ (\leftrightarrow e \setminus A \neq \emptyset).$$

The set of all boundary points of A is denoted $\mathbf{Fr}(\mathbf{A})$.

From its definition,

$$\mathbf{Fr}(\mathbf{A}) = \mathbf{Fr}(\mathbf{A^c}).$$

Closure, Interior, and Exterior

With the boundary operator **Fr**, we can define

Closure

- Closure $\mathbf{Cl}(\mathbf{A}) = \mathbf{A} \cup \mathbf{Fr}(\mathbf{A})$;
- Interior $\mathbf{Int}(\mathbf{A}) = \mathbf{A} \setminus \mathbf{Fr}(\mathbf{A})$;
- Exterior $\mathbf{Ext}(\mathbf{A}) = \mathbf{V} \setminus \mathbf{Cl}(\mathbf{A})$.

Therefore,

- $\mathbf{Cl}(\mathbf{A}) = \mathbf{Int}(\mathbf{A}) \cup \mathbf{Fr}(\mathbf{A})$;
- $\mathbf{Int}(\mathbf{A}) \subseteq \mathbf{A} \subseteq \mathbf{Cl}(\mathbf{A})$;
- Set V is decomposed into three non-intersecting parts:

$$V = \mathbf{Int}(\mathbf{A}) \cup \mathbf{Fr}(\mathbf{A}) \cup \mathbf{Ext}(\mathbf{A}).$$

Closure, Interior, and Exterior

With the boundary operator **Fr**, we can define

Closure

- Closure $\mathbf{Cl}(\mathbf{A}) = \mathbf{A} \cup \mathbf{Fr}(\mathbf{A})$;
- Interior $\mathbf{Int}(\mathbf{A}) = \mathbf{A} \setminus \mathbf{Fr}(\mathbf{A})$;
- Exterior $\mathbf{Ext}(\mathbf{A}) = \mathbf{V} \setminus \mathbf{Cl}(\mathbf{A})$.

Therefore,

- $\mathbf{Cl}(\mathbf{A}) = \mathbf{Int}(\mathbf{A}) \cup \mathbf{Fr}(\mathbf{A})$;
- $\mathbf{Int}(\mathbf{A}) \subseteq \mathbf{A} \subseteq \mathbf{Cl}(\mathbf{A})$;
- Set V is decomposed into three non-intersecting parts:

$$V = \mathbf{Int}(\mathbf{A}) \cup \mathbf{Fr}(\mathbf{A}) \cup \mathbf{Ext}(\mathbf{A}).$$

Accumulation Point

Given a cross-table (V, E), we define two types of points with respect to any non-empty subset $A \subseteq V$ (which does not have to be a member of E):

Accumulation Point

A point $v \in V$ (not necessarily in A) is called A's *accumulation point*, if for every $e \in E(v)$,

$$e \cap A \setminus \{v\} \neq \emptyset,$$

that is, every neighbor of the point v contains at least one point of A other than v.

- An accumulation point v of A may not be in A;
- Any accumulation point v can be "approached" via a sequence of points in A, each of which is in a neighborhood $e \in E(v)$ of v. We can construct a learning

Derived set and Isolation Set

The set of all accumulation points of A is called the *derived set* of A, denoted $\mathbf{Der(A)}$. Any point of A which is not an accumulation point is called *isolation point*; the set of isolated points of A is denoted $\mathbf{Iso(A)}$.

Theorem: Two Partition of Closed Sets

- $\mathbf{Cl(A)} = \mathbf{Int(A)} \cup \mathbf{Fr(A)}$;
- $\mathbf{Cl(A)} = \mathbf{Iso(A)} \cup \mathbf{Der(A)}$;

which means

$$\mathbf{Cl(A)} = \mathbf{A} \cup (\mathbf{A^c} \cap \mathbf{Fr(A)} \cap \mathbf{Der(A)});$$

Derived set and Isolation Set

The set of all accumulation points of A is called the *derived set* of A, denoted $\mathbf{Der(A)}$. Any point of A which is not an accumulation point is called *isolation point*; the set of isolated points of A is denoted $\mathbf{Iso(A)}$.

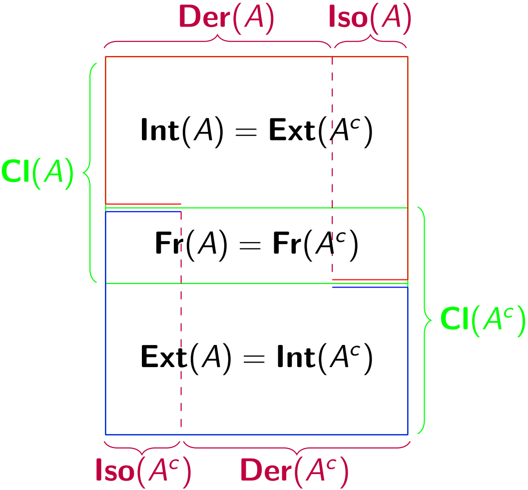
Theorem: Two Partition of Closed Sets

- $\mathbf{Cl(A)} = \mathbf{Int(A)} \cup \mathbf{Fr(A)}$;
- $\mathbf{Cl(A)} = \mathbf{Iso(A)} \cup \mathbf{Der(A)}$;

which means

$$\mathbf{Cl(A)} = \mathbf{A} \cup (\mathbf{A^c} \cap \mathbf{Fr(A)} \cap \mathbf{Der(A)});$$

Main Result: Partition of V through arbitrary $A \subset V$



Thank You!

- Zhang, H., Xu, Y., and Zhang, J. (2009) Reproducing kernel Banach spaces for machine learning. *Journal of Machine Learning Research.* 10: 2741-2775.
- Zhang, H. and Zhang, J. (2010). Generalized semi-inner products with application to regularized learning. *Journal of Mathematical Analysis and Application.* 372: 181-196.
- Zhang, H. and Zhang, J. (2011). Frames, Riesz bases, and sampling expansions in Banach spaces via semi-inner products. *Applied and Computational Harmonic Analysis.* 31:1-25.
- Zhang, H. and Zhang, J. (2012). Regularized learning in Banach space as an optimization problem: Representer theorems. *Journal of Global Optimization.* 54: 235-250
- Zhang, H. and Zhang, J. (2013). Vector-valued Reproducing Kernel Banach Spaces with applications to multi-task learning, *Journal of Complexity.* 29: 195-215.
- Zhang, H. and Zhang, J. (2015). Learning with Reproducing Kernel Banach Spaces. *Proceedings of the 10th ISAAC Congress, Macau.* In *New Trends in Analysis and Interdisciplinary Applications.* Birkhauser.
- Zhang, H. and Zhang, J. (2017). Learning with reproducing kernel Banach spaces. New Trends in Analysis and Interdisciplinary Applications. Springer.