Learning with Invariance via Linear Functionals on RKHS

Xinhua Zhang[1], Wee Sun Lee[2], Yee Whye Teh[3]



[1] Machine Learning Research Group, NICTA and Australian National University

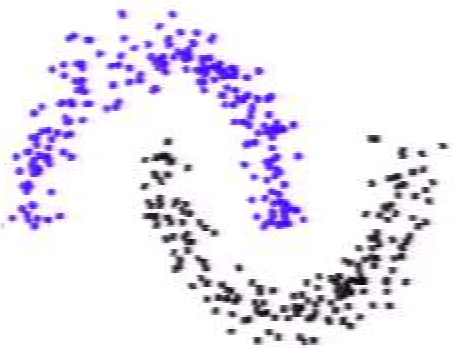
[2] Department of Computer Science, National University of Singapore

[3] Department of Statistics, University College, University of Oxford

Invariance represented as linear functional

- Invariance: target function f keeps invariant under prescribed transforms
- Goal 1: *compactly* encode *varieties* of invariance for a *rich* family of function
- Goal 2: efficient search of the optimal function via convex programming
- Ex 1: graph Laplacian: smooth change on data manifold

$$\sum_{i\sim j}[f(x_i)-f(x_j)]^2$$



- Ex 2: differentiation: flat gradient at observed data x_i

$$\frac{\partial}{\partial x^d}\bigg|_{x=x_i} f(x)$$

- Ex 3: invariance under transformation $T(x;\theta)$, *eg*. T for rotation, θ for degree
  - Small $f(T(x_i;\theta)) - f(T(x_i;0))$
  - Approximation by gradient $\frac{\partial}{\partial \theta}\bigg|_{\theta=0} f(T(x,\theta))$

- Ex 4: local average $\int_{\mathcal{X}} f(\tau)p(x_i-\tau)\mathrm{d}\tau - f(x_i)$

Bounded linear functional and Representer theorem

- Functional interpretation $L : \mathcal{H} \to \mathbb{R}$
- Linearity: all above invariance scores are linear functionals in f
- Bounded for a number of combinations of $\mathcal{H}$ and L
  - Riesz representation theorem in RKHS: any L can be represented as

$$z \in \mathcal{H} \quad \text{such that} \quad L(f) = \langle f, z \rangle \quad \forall f \in \mathcal{H}$$

- Regularized risk minimization

$$\min_{f \in \mathcal{H}} \quad \frac{1}{2}\|f\|^2 + \lambda \sum_{i=1}^{l} \ell_1(f(x_i), y_i) + \nu \sum_{j=1}^{m} \ell_2 (L_j(f))$$

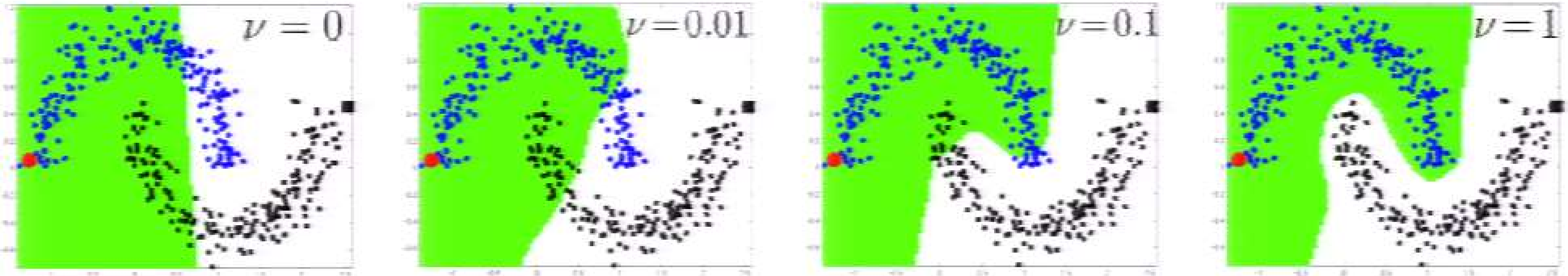
  - Representer theorem

$$f^*(\cdot) = \sum_{i=1}^{l} \alpha_i k(x_i, \cdot) + \sum_{j=1}^{m} \beta_j z_j(\cdot)$$

  - Efficient convex optimization: no expensive SDP, SOCP
  - Compact representation of invariance using kernels: no need of virtual samples

Experimental results
(more results at the poster)

Semi-supervised learning: gradient invariance



Transformation invariance: shift, rotation, scaling, and shearing (MNIST)