Reproducing Kernel Hilbert Space

Let $X \subseteq \mathbb{R}^p$ and $H = \{f : f : \mathcal{X} \to \mathbb{R}\}$ s.t.

1. The collection forms a V.S.
2. Has an inner product $\langle \cdot, \cdot \rangle_H$
3. Complete
4. For each $x \in \mathcal{X}$ $f(x) = \langle f(\cdot), K(\cdot, x) \rangle_H$ $\forall f \in H$.

   and some Kernel function $K(\cdot, \cdot)$.

Then H is a reproducing Kernel Hilbert Space.

Thrm: (Representer theorm)

$$\hat{f} = \underset{f \in H}{\arg\inf} \left\{ \sum_{i=1}^{n} L(f(x_i), y_i) + \lambda \|f\|_H \right\}$$

$$= \sum_{i=1}^{n} \alpha_i K(\cdot, x_i)$$

Calculate $\hat{f}(x_i) = \sum_{j=1}^{n} \alpha_j K(x_i, x_j) \equiv (K\alpha)_i$

Now taking the loss to be ℓ_2 then the problem becomes

$$\hat{f} = \arg\min_{\alpha} \left\{ (K\alpha - Y)^T (K\alpha - Y) + \lambda \|f\|_{\mathcal{H}}^2 \right\}$$

$$\|f\|_{\mathcal{H}} = \sqrt{\sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j K(x_i, x_j)}$$

So the optimal α

$$\hat{\alpha} = \arg\min_{\alpha} (K\alpha - Y)^T (K\alpha - Y) + \lambda \alpha^T K \alpha$$

$$= 2K^T (Y - K\hat{\alpha}) + 2\lambda \alpha = 0$$

$$\Longrightarrow \hat{\alpha} = (K^T K + \lambda K)^{-1} K^T Y$$

as K is p.s.d./symmetric

$$= (K(K + \lambda))^{-1} K Y$$

$$\boxed{= (K + \lambda I)^{-1} Y}$$

Kernel Logistic Regression

$$\log \mathbb{P}(Y = 1) = f(x) \in \mathcal{H}$$

$$\overline{\mathbb{P}(Y=0)}$$

The function we look to minimize is given by

$$\mathcal{L} = \frac{1}{n}\sum_{i=1}^{n} \log\left\{1+\exp\{-y_i f(x_i)\}\right\} + \lambda \|f\|_{\mathcal{H}}^2$$

$\hat{f}(\cdot) = \sum_{i=1}^{n} \alpha_i K(\cdot, x_i)$ then use N.R. to get estimates for α_i.

<u>Ex:</u> Take $\mathcal{H} = \{f : \int (\Delta f - f)^2 \, dx < \infty\}$

$K(y, x) = \mathcal{F}^{-1}\left(\frac{1}{\|w\|^2+1}\right)$ w from $(\Delta - I) f = w$

Support Vector Machine

Let's use Hinge Loss and define $K(x,y) = x \cdot y + 1$

$$\mathcal{H} = \{f = x^T\beta + \beta_0\}$$

$f_1(x) = x^T\beta + \beta_0 \qquad f_2(x) = x^T\tilde{\beta} + \tilde{\beta}_0$

$$\langle f_1, f_2 \rangle = \beta^T\tilde{\beta} + \beta_0\tilde{\beta}_0$$

$$\hat{f}(x) = \underset{\mathcal{H}}{\operatorname{argmin}} \sum_{i=1}^{n} \left(1 - y_i f(x_i)\right)_+ + \lambda \|f\|_{\mathcal{H}}^2$$

$$= \sum_{i=1}^{n} \alpha_i \, K(x, x_i) = \sum_{i=1}^{n} \alpha_i \left(x_i^T x + 1 \right)$$

$$= x^T \underbrace{\left[\sum_{i=1}^{n} (\alpha_i x_i^T) \right]}_{\beta} + \underbrace{\sum_{i=1}^{n} \alpha_i}_{\beta_0}$$

So plugging back into our loss

$$\hat{f} = x^T \vec{\beta} + \vec{\beta}_0 = \underset{(\beta_0, \beta)}{\operatorname{arginf}} \left\{ \sum_{i=1}^{n} \left[1 - y_i (x^T \beta + \beta_0) \right]_+ + \lambda \left(\|\beta\|^2 + \beta_0^2 \right) \right\}$$

Expand space of Kernels $K(x, y) = (x^T y + 1)^d$

$\mathcal{H}_d = \{$ polynomials in x of order d $\}$