KERNEL RIDGE REGRESSION

Recall ridge regression: Given $(x_i, y_i)_{i=1}^n$, $x_i \in \mathbb{R}^d$, $y_i \in \mathbb{R}$

$$(\hat{\beta}, \hat{\beta}_0) \leftarrow \min_{\beta, \beta_0} \sum_{i=1}^n (y_i - \beta^T x_i - \beta_0)^2 + \lambda \|\beta\|^2$$

Solution:

$$\left.\begin{aligned} \hat{\beta} &= (A^T A + \lambda I)^{-1} A^T \tilde{y} \\ \hat{\beta}_0 &= \bar{y} - \hat{\beta}^T \bar{x} \end{aligned}\right\} \Rightarrow \begin{aligned} \hat{f}(x) &= \hat{\beta}^T x + \beta_0 \\ &= \bar{y} + \hat{\beta}^T (x - \bar{x}) \end{aligned}$$

where

$$A = \begin{bmatrix} \tilde{x}_1^T \\ \vdots \\ \tilde{x}_n^T \end{bmatrix}, \quad \begin{aligned} \tilde{x}_i &= x_i - \bar{x} \\ \tilde{y}_i &= y_i - \bar{y} \end{aligned}, \quad \tilde{y} = \begin{bmatrix} \tilde{y}_1 \\ \vdots \\ \tilde{y}_n \end{bmatrix}$$

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \qquad \bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$$

Can we express RR in terms of inner products $\langle x_i, x_j \rangle$ and $\langle x_i, x \rangle$?

Not immediately. Note $A^T A$ is not $[\langle \tilde{x}_i, \tilde{x}_j \rangle]_{i,j=1}^n$

Let's apply the matrix inversion lemma, also known as the Woodbury matrix identity:

$$(P + QRS)^{-1} = P^{-1} - P^{-1}Q\,(R^{-1} + SP^{-1}Q)^{-1}SP^{-1}$$

where

$$P = \lambda I, \quad Q = A^T, \quad R = I, \quad S = A.$$

We have

$$(\lambda I + A^T A)^{-1} = \frac{1}{\lambda} I - \frac{1}{\lambda} I \cdot A^T \left(I + \frac{1}{\lambda} A A^T\right)^{-1} A \cdot \frac{1}{\lambda}$$

$$= \frac{1}{\lambda}\left[I - A^T(\lambda I + AA^T)^{-1}A\right]$$

$$\Longrightarrow \quad (A^T A + \lambda I)^{-1} A^T \tilde{y} = \frac{1}{\lambda}\left[A^T - A^T(A A^T + \lambda I)^{-1} A A^T\right]\tilde{y}$$

$$= \frac{1}{\lambda}\left[A^T - A^T(K + \lambda I)^{-1} K\right]\tilde{y}$$

where $K = [\langle \tilde{x}_i, \tilde{x}_j \rangle]_{i,j=1}^n$. Note $\langle \tilde{x}_i, \tilde{x}_j \rangle$

$$= \langle x_i, x_j \rangle - \frac{1}{n}\sum_{r=1}^n \langle x_i, x_r \rangle - \frac{1}{n}\sum_{s=1}^n \langle x_s, x_j \rangle + \frac{1}{n^2}\sum_{r,s=1}^n \langle x_r, x_s \rangle.$$

What about the remaining A^T?

$$\hat{f}(x) = \bar{y} + \hat{\beta}^T (x - \bar{x})$$

$$= \bar{y} + \frac{1}{\lambda} \tilde{y} \left[A - K (K + \lambda I)^{-1} A \right] (x - \bar{x})$$

$$= \bar{y} + \frac{1}{\lambda} \tilde{y} \left[I - K (K + \lambda I)^{-1} \right] \underline{k}(x)$$

where

$$\underline{k}(x) = \begin{bmatrix} \langle \tilde{x}_1, x - \bar{x} \rangle \\ \vdots \\ \langle \tilde{x}_n, x - \bar{x} \rangle \end{bmatrix}$$

Note:

$$\langle \tilde{x}_i, x - \bar{x} \rangle$$

$$= \langle x_i - \bar{x}, x - \bar{x} \rangle$$

$$= \langle x_i, x \rangle - \frac{1}{n} \sum_j \langle x_i, x_j \rangle$$

$$- \frac{1}{n} \sum_j \langle x, x_j \rangle + \frac{1}{n} \sum_{j,k} \langle x_j, x_k \rangle$$

Also note

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

(A) $=$

$=$

$$\Longrightarrow \hat{f}(x) = \bar{y} + \tilde{y} (K + \lambda I)^{-1} \underline{k}(x)$$

For some kernels, $k(x,x') = \langle \Phi(x), \Phi(x') \rangle$, $\Phi(x)$ already contains a constant component, in which case β_0 is not needed. An example is the inhomogeneous polynomial kernels.

The Gaussian kernel also does not seem to require β_0.

If β_0 is omitted, the KRR solution is

$$\hat{f}(x) = \tilde{y}(K+\lambda I)^{-1}\underline{k}(x)$$

where

(5) $K =$ $\underline{k}(x) =$

Example | Gaussian kernel

$$k(x, x') = \exp\left\{ -\frac{\|x - x'\|^2}{2\sigma^2} \right\}$$

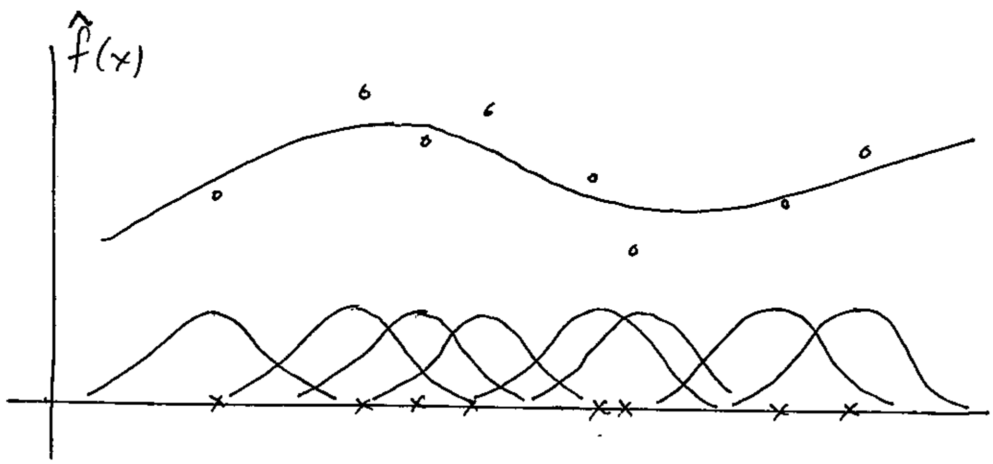
Then

$$\hat{\beta}^T x = \tilde{\underline{y}}^T (K + \lambda I)^{-1} \underline{k}(x)$$

$$= \alpha^T \underline{k}(x)$$

$\boxed{\underline{\alpha} \text{ independent of } x}$

$$= \sum \alpha_i \, k(x, x_i)$$



$\boxed{\text{Key}}$ A.

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

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

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

B.

$$K = \begin{bmatrix} k(x_1, x_1) & \cdots & k(x_1, x_n) \\ \vdots & \ddots & \\ k(x_n, x_1) & \cdots & k(x_n, x_n) \end{bmatrix}, \quad \underline{k}(x) = \begin{bmatrix} k(x, x_1) \\ \vdots \\ k(x, x_n) \end{bmatrix}$$