

Assessment_2

2023-03-15

Kernel Methods For Regression

Part 1

Question 1

The Gaussian kernel is an example of a kernel for which the model is identifiable. The model is unidentifiable for the kernel: $k(x, y) = 1$ for $x, y \in \mathbb{R}^p$. This kernel is positive semi-definite and by the Moore-Aronszajn theorem there exists a unique RKHS for which k is the reproducing kernel. In this RKHS all the functions are constant, let's pick two functions $f_1, f_2 \in H_k$ where $f_1(x) = c, f_2(x) = d$ for all $x \in \mathcal{X}$. Then if we let the α we use with f_2 , $\alpha_2 = \alpha_1 - d + c$, where α_1 is the α we use with f_1 , then these are the same model. Hence our model is unidentifiable.

Question 2

We have for some $f \in H_k$: $f = f_1 + f_2$, for some $f_1 \in \tilde{H}, f_2 \in \tilde{H}^\perp$. By orthogonality:

$$\|f_1 + f_2\|^2 = \|f_1\|^2 + \|f_2\|^2$$

And by the reproducing property:

$$\begin{aligned} & \frac{1}{2n} \sum_{i=1}^n \log f(y_i; g^{-1}(\alpha + (f_1 + f_2)(x_i^0)), \phi) \\ &= \frac{1}{2n} \sum_{i=1}^n \log f(y_i; g^{-1}(\alpha + f_1(x_i^0)), \phi) \end{aligned}$$

Combining the above,

$$\begin{aligned} & \frac{1}{2n} \sum_{i=1}^n \log f(y_i; g^{-1}(\alpha + (f_1 + f_2)(x_i^0)), \phi) - \lambda \|f_1 + f_2\|^2 \\ &= \frac{1}{2n} \sum_{i=1}^n \log f(y_i; g^{-1}(\alpha + f_1(x_i^0)), \phi) - \lambda \|f_1\|^2 + \lambda \|f_2\|^2 \\ &\geq \frac{1}{2n} \sum_{i=1}^n \log f(y_i; g^{-1}(\alpha + f_1(x_i^0)), \phi) - \lambda \|f_1\|^2 \end{aligned}$$

Hence, we have that $\hat{f}_\lambda \in \tilde{H}$ and therefore can write \hat{f}_λ as a linear combination of $k(x_1^0, \cdot), \dots, k(x_n^0, \cdot)$ therefore we can write:

$$\hat{f}_\lambda = \sum_{i=1}^n \hat{\beta}_{\lambda, i} k(x_i^0, \cdot)$$

Question 3

We have that,

$$\begin{aligned} -\lambda \|f\|_{H_k}^2 &= -\lambda \left\| \sum_{i=1}^n \beta_{\lambda,i} k(x_i^0, \cdot) \right\|_{H_k}^2 \\ &= -\lambda \left\langle \sum_{i=1}^n \beta_{\lambda,i} k(x_i^0, \cdot), \sum_{j=1}^n \beta_{\lambda,j} k(x_j^0, \cdot) \right\rangle_k = -\lambda \sum_{i=1}^n \beta_{\lambda,i} \left\langle k(x_i^0, \cdot), \sum_{j=1}^n \beta_{\lambda,j} k(x_j^0, \cdot) \right\rangle_k \\ &= -\lambda \sum_{i=1}^n \sum_{j=1}^n \beta_{\lambda,i} \beta_{\lambda,j} \langle k(x_i^0, \cdot), k(x_j^0, \cdot) \rangle_k \end{aligned}$$

Where $K = [k(x_i^0, x_j^0)]_{i,j}$. Hence we can rewrite (2) as:

$$\frac{1}{2n} \sum_{i=1}^n \log f(y_i; g^{-1}(\alpha + \beta_{\lambda}^T \cdot K^{(i)}), \phi) - \lambda \beta_{\lambda}^T K \beta_{\lambda}$$

Question 4

Let $m < n + 2$, consider the optimization problem posed in the previous question, the Nyonstrom method involves reducing the dimension of the gram matrix, K , hence reducing the dimensionality of the optimization problem. We will approximate the kernel, k , by $\tilde{k}^{(m)}$ such that the matrix $\tilde{K}^{(m)}$ obtained by replacing k with $\tilde{k}^{(m)}$ has rank $\leq m$. We will let,

$$\tilde{k}^{(m)}(x, x') = k_m(x)^T (K_m)^{-1} k_m(x')$$

where K_m is the first m rows and columns of K and

$$k_m(x) = (k(x_1^0, x), \dots, k(x_m^0, x))$$

hence leaving the following optimization problem,

$$\arg \max_{\alpha \in \mathbb{R}, \phi \in (0, \inf), \beta'_{\lambda} \in \mathbb{R}^m} \frac{1}{2n} \sum_{i=1}^n \log f(y_i; g^{-1}(\alpha + \beta_{\lambda}'^T \cdot K^{(i)}), \phi) - \lambda \beta_{\lambda}'^T \tilde{K}^{(m)} \beta_{\lambda}'$$

where we write β'_{λ} for our now lower dimensional β_{λ} .

Question 5

Let's first find the spectral decomposition of $\tilde{K}^{(m)}$,

$$\tilde{K}^{(m)} = S \Lambda S^{-1}$$

Then we can write,

$$\beta_{\lambda}'^T \tilde{K}^{(m)} \beta_{\lambda}' = \beta_{\lambda}'^T S^{-T} \Lambda S^{-1} \beta_{\lambda}' = \|\Lambda^{\frac{1}{2}} S^{-1} \beta_{\lambda}'\|_2^2$$

Glmnet estimates parameters that minimize the following:

$$-\log f(y_i; g^{-1}(\alpha + \beta_{glm}^T K^{(i)}), \phi) + \lambda \|\beta_{glm}\|_2^2$$

which is the same as maximizing optimization problem as ours, except for if we instead try to find the minimum of the negative of our optimization problem and if we substitute for β_{glm} . We have that,

$$\begin{aligned} \beta_{glm} &= \Lambda^{\frac{1}{2}} S \beta_{\lambda}' \\ \Rightarrow \beta_{\lambda}' &= S^{-1} \Lambda^{-\frac{1}{2}} \beta_{glm} \end{aligned}$$

Therefore we can find our value for β'_{λ} using the estimate we get from glmnet.

Part 2

Question 6