ASSIGNMENT SHEET 11

Nayan Man Singh Pradhan

Excercise 1. (Polynomial models)

In this paper & pencil task, you compare the quadratic model based on the basis functions $\{1, X, X^2\}$ to the linear model for the following training set.

$$\mathcal{T} = \{(-1, 1), (0, 0.5), (1, 3)\}.$$

a) Compute the linear model using least squares for the given training data.

b) Compute the quadratic model using least squares for the given training data.

c) Draw (by hand) the linear and the quadratic model in a plot, together with the training samples.

(4 Points)

a) $X = \begin{pmatrix} 1 & -1 \\ 1 & 0 \\ 1 & 1 \end{pmatrix}, \quad X^T = \begin{pmatrix} 1 & 1 & 1 \\ -1 & 0 & 1 \end{pmatrix}, \quad y = \begin{pmatrix} 1 \\ 0.5 \\ 3 \end{pmatrix}$

$$X^T X = \begin{pmatrix} 1 & 1 & 1 \\ -1 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & -1 \\ 1 & 0 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} 3 & 0 \\ 0 & 2 \end{pmatrix}$$

$$X^T y = \begin{pmatrix} 1 & 1 & 1 \\ -1 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 0.5 \\ 3 \end{pmatrix} = \begin{pmatrix} 4.5 \\ 2 \end{pmatrix}$$

$$X^T X \hat{\beta} = X^T y$$

$$\begin{pmatrix} 3 & 0 \\ 0 & 2 \end{pmatrix} \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix} = \begin{pmatrix} 4.5 \\ 2 \end{pmatrix}$$

$\hat{\beta}_1 = 1.5 \qquad \hat{\beta}_2 = 1$

$Y = 1.5 + X$ //

$\therefore$ The linear model using least squares for the given training data is: $Y_L = 1.5 + X$

b) $\varphi_1(x) = 1$

$\varphi_2(x) = x$

$\varphi_3(x) = x^2$

$$X = \begin{pmatrix} \varphi_1(x_1) & \varphi_2(x_1) & \varphi_3(x_1) \\ \varphi_1(x_2) & \varphi_2(x_2) & \varphi_3(x_2) \\ \varphi_1(x_3) & \varphi_2(x_3) & \varphi_3(x_3) \end{pmatrix} = \begin{pmatrix} 1 & -1 & 1 \\ 1 & 0 & 0 \\ 1 & 1 & 1 \end{pmatrix}$$

$$X^T = \begin{pmatrix} 1 & 1 & 1 \\ -1 & 0 & 1 \\ 1 & 0 & 1 \end{pmatrix}$$

$$X^T X = \begin{pmatrix} 1 & 1 & 1 \\ -1 & 0 & 1 \\ 1 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & -1 & 1 \\ 1 & 0 & 0 \\ 1 & 1 & 1 \end{pmatrix} = \begin{pmatrix} 3 & 0 & 2 \\ 0 & 2 & 0 \\ 2 & 0 & 2 \end{pmatrix}$$

$$X^T y = \begin{pmatrix} 1 & 1 & 1 \\ -1 & 0 & 1 \\ 1 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 0.5 \\ 3 \end{pmatrix} = \begin{pmatrix} 4.5 \\ 2 \\ 4 \end{pmatrix}$$

$$X^T X \hat{\alpha} = X^T y$$

$$\begin{pmatrix} 3 & 0 & 2 \\ 0 & 2 & 0 \\ 2 & 0 & 2 \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{pmatrix} = \begin{pmatrix} 4.5 \\ 2 \\ 4 \end{pmatrix}$$

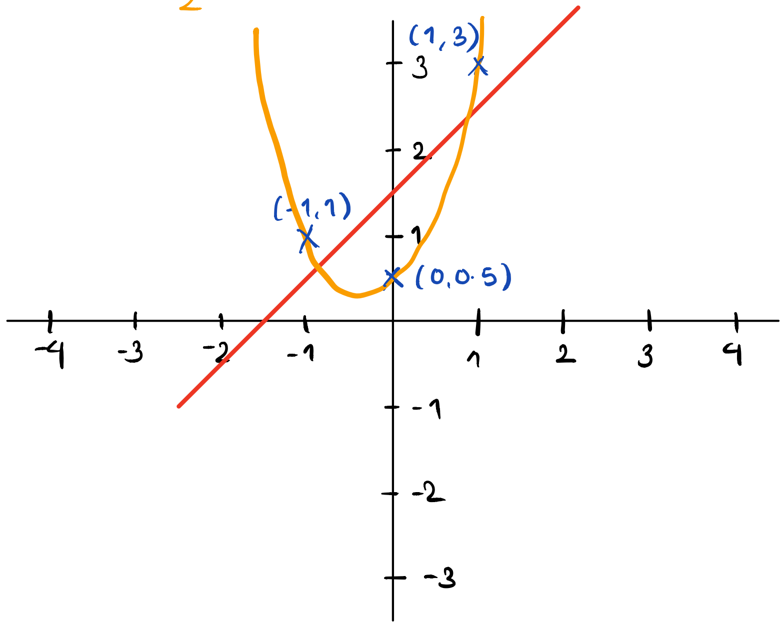
$\alpha_1 = \frac{1}{2}$ $\qquad \alpha_2 = 1$ $\qquad \alpha_3 = \frac{3}{2}$

$$Y = \frac{1}{2} + x + \frac{3}{2} x^2$$

$\therefore$ The quad. model using least squares for the given training data is: $Y_Q = \frac{1}{2} + x + \frac{3}{2} x^2$

c)

$$Y_Q = \frac{1}{2} + X + \frac{3}{2}X^2 \qquad Y_L = 1.5 + X$$



Excercise 2. (Kernel-based model)

You are given the training data

$$\mathcal{T} = \{(-2,2)^\top, (0,0)^\top, (1,2)^\top\}.$$

a) Use the Gaussian kernel with kernel width $\sigma = 1$ and compute by hand the kernel-based model using least squares for the given training data.

b) Repeat the previous task, but this time you compute by hand the kernel-based model using ridge regression with regularization parameter $\lambda = 1$.

c) Draw both above models including the training data in one plot.

(4 Points)

a) We know

$$k(x,x') = \exp\left(-\frac{\|x-x'\|_2^2}{\sigma^2}\right)$$

$$A = \begin{pmatrix} k(x_1,x_1) & k(x_1,x_2) & k(x_1,x_3) \\ k(x_2,x_1) & k(x_2,x_2) & k(x_2,x_3) \\ k(x_3,x_1) & k(x_3,x_2) & k(x_3,x_3) \end{pmatrix}$$

$$= \begin{pmatrix} k(-2,-2) & k(-2,0) & k(-2,1) \\ k(0,-2) & k(0,0) & k(0,1) \\ k(1,-2) & k(1,0) & k(1,1) \end{pmatrix}$$

$$= \begin{pmatrix} 1 & 0.0183 & 1.234\times10^{-4} \\ 0.0183 & 1 & 0.36787 \\ 1.234\times10^{-4} & 0.36787 & 1 \end{pmatrix}$$

Now, $A\hat{\alpha} = y$

$$\begin{pmatrix} 1 & 0.0183 & 1.234\times10^{-4} \\ 0.0183 & 1 & 0.36787 \\ 1.234\times10^{-4} & 0.36787 & 1 \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{pmatrix} = \begin{pmatrix} 2 \\ 0 \\ 2 \end{pmatrix}$$

$\alpha_1 = 2.02$, $\alpha_2 = -0.893$, $\alpha_3 = 2.33$

$$Y_{ker} = 2.02\, k(x,x_0) - 0.893\, k(x,x_1) + 2.33\, k(x,x_2)$$

$$\therefore Y_{ker} = 2.02\, e^{-(\|x+2\|^2)} - 0.893\, e^{-(\|x\|^2)} + 2.33\, e^{-(\|x-1\|^2)}$$

b) Kernel based ridge regression.

From previous exercise

$$A = \begin{pmatrix} 1 & 0.0183 & 1.234\times10^{-4} \\ 0.0183 & 1 & 0.36787 \\ 1.234\times10^{-4} & 0.36787 & 1 \end{pmatrix}$$

$$A+\lambda I = \begin{pmatrix} 1 & 0.0183 & 1.234\times10^{-4} \\ 0.0183 & 1 & 0.36787 \\ 1.234\times10^{-4} & 0.36787 & 1 \end{pmatrix} + \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}_{\lambda=1}$$

$$= \begin{pmatrix} 2 & 0.0183 & 1.234\times10^{-4} \\ 0.0183 & 2 & 0.36787 \\ 1.234\times10^{-4} & 0.36787 & 2 \end{pmatrix}$$

Now,

$$(A+\lambda I)\,\hat{\alpha} = y$$

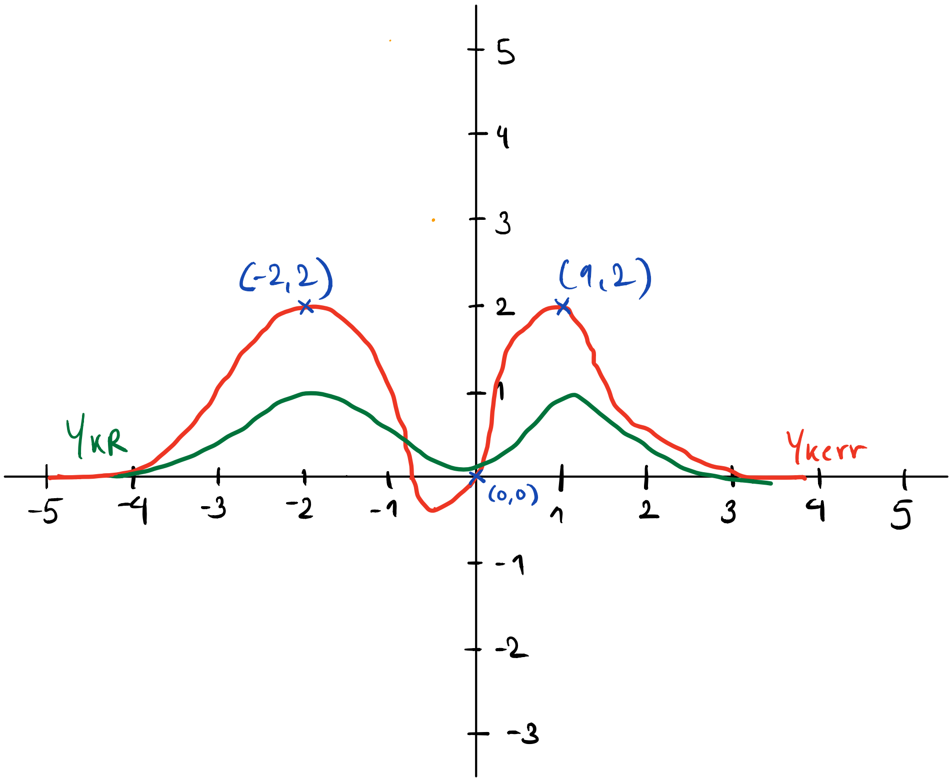
$$\begin{pmatrix} 2 & 0.0183 & 1.234\times10^{-4} \\ 0.0183 & 2 & 0.36787 \\ 1.234\times10^{-4} & 0.36787 & 2 \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{pmatrix} = \begin{pmatrix} 2 \\ 0 \\ 2 \end{pmatrix}$$

$$\alpha_1 = 1.00 \quad , \alpha_2 = -0.200 \, , \alpha_3 = 1.04$$

$$Y_{KR} = 1\ k(x,x_0) - 0.200\, k(x,x_1) + 1.04\ k(x,x_2)$$

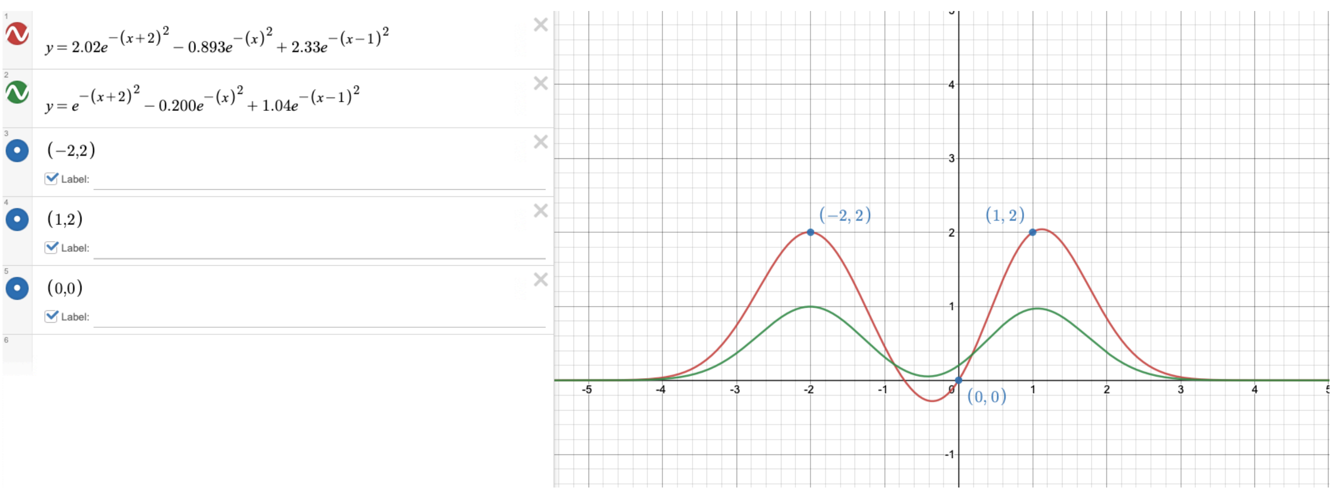
$$Y_{KR} = e^{-(\|x+2\|^2)} - 0.200\, e^{-(\|x\|^2)} + 1.04\, e^{-(\|x-1\|^2)}$$ //

c)



$$y_{ker} = 2.02\, e^{-(\|x+2\|^2)} - 0.893\, e^{-(\|x\|^2)} + 2.33\, e^{-(\|x-1\|^2)}$$

$$y_{KR} = e^{-(\|x+2\|^2)} - 0.200\, e^{-(\|x\|^2)} + 1.04\, e^{-(\|x-1\|^2)}$$



Excercise 3. (Kernel matrix)

Prove Lemma 10.2 from the lecture notes.

(4 Points)

Lemma 10.2 Let $\{\boldsymbol{x}_i\}_{i=1}^N$ be a set of points such that $\boldsymbol{x}_i \in \mathbb{R}^D$ and let $k : \mathbb{R}^D \times \mathbb{R}^D \to \mathbb{R}$ be a positive definite kernel. Then, the matrix

$$\mathcal{A} = \begin{pmatrix} k(\boldsymbol{x}_1, \boldsymbol{x}_1) & \cdots & k(\boldsymbol{x}_1, \boldsymbol{x}_N) \\ \vdots & \ddots & \vdots \\ k(\boldsymbol{x}_N, \boldsymbol{x}_1) & \cdots & k(\boldsymbol{x}_N, \boldsymbol{x}_N) \end{pmatrix}$$

is (symmetric) positive definite.

From Definition 10.3,

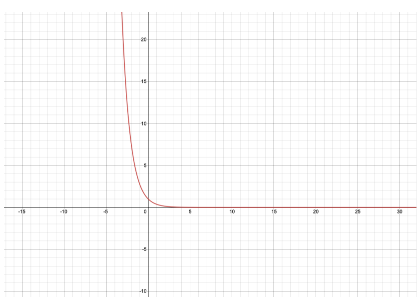
A continuous bivariate function $k : \mathbb{R}^D \times \mathbb{R}^D \to \mathbb{R}$ is called (symmetric) positive definite kernel on $\mathbb{R}^D$, if for all $N \in \mathbb{N}$, all finite subsets $\{x_1, \dots, x_N\} \subseteq \mathbb{R}^D$ and all $\alpha \in \mathbb{R}^N \setminus \{0\}$, we have

$$\sum_{j=1}^{N} \sum_{\ell=1}^{N} \alpha_j \alpha_\ell k(x_j, x_\ell) \geq 0$$

'A' is a symmetric matrix as the kernel outputs the same value for $k(x_i, x_j)$ and $k(x_j, x_i)$ for two samples 'i' and 'j'.

'A' is a positive definite matrix as the kernel function k that makes up A is positive definite. In order for a matrix to be positive definite, all eigenvalues of the matrix should be positive. Alternatively, the Sylvester's criterion should be fulfilled.

First, we establish that the value of Gaussian kernel $k(x, x') = e^{-\frac{(\|x-x'\|_2^2)}{\sigma^2}}$, will always be positive as $\|x-x'\|_2^2 \geq 0$ and $\sigma^2 > 0$ and $e^{-a} \geq 0$ if $a \geq 0$ (as can be seen in the graph)



According to Sylvester's criterion,

$$\det(k(x_1,x_1)) = k(x_1,x_1) > 0$$

$$\det\begin{pmatrix} k(x_1,x_1) & k(x_1,x_2) \\ k(x_2,x_1) & k(x_2,x_2) \end{pmatrix} = [k(x_1,x_1)\cdot k(x_2,x_2) - k(x_1,x_2)^2] > 0$$

$$\vdots$$

and so on.

Hence it proves that A is positive definite!

$\therefore$ Lemma 10.2 Proved!