

“Machine Learning and Computational Statistics”

2nd Homework

Exercise 1:

Consider the following nonlinear model:

$$y = 3x_1^2 + 4x_2^2 + 5x_3^2 + 7x_1x_2 + x_1x_3 + 4x_2x_3 - 2x_1 - 3x_2 - 5x_3 + \eta$$

Define a suitable function φ that transforms the problem to a space where the problem of estimating the model becomes linear. What is the dimension of the original and the transformed spaces?

Exercise 2:

Consider the following two-class nonlinear classification task:

$$\mathbf{x} = [x_1, x_2, x_3]^T: x_1^2 + 3x_2^2 + 6x_3^2 + x_1x_2 + x_2x_3 > (<)3 \rightarrow \mathbf{x} \in \omega_1(\omega_2)$$

Define a suitable function φ that transforms the problem to a space where the problem of estimating the border of the two classes becomes linear. What is the dimension of the original and the transformed spaces?

Exercise 3 (python code + text):

- (a) **Generate** a set $\mathbf{X} = \{(y_i, \mathbf{x}_i), \mathbf{x}_i = [x_{i1}, x_{i2}]^T \in \mathbb{R}^2, y_i \in \mathbb{R}, i = 1, \dots, 200\}$ from the model

$$y = \theta_0 + \theta_1x_1 + \theta_2x_2 + \theta_3x_1x_2 + \eta$$

where η is an i.i.d. normal zero mean noise, with variance 0.05. Use $\theta_0 = 3, \theta_1 = 2, \theta_2 = 1, \theta_3 = 1$ (adopt the strategy given in the example of the 2nd slide of the 2nd lecture). In the sequel, pretend that you do not know the model that generates the data. All you have at your disposal is the data set \mathbf{X} .

- (b) **Adopting** the linear model assumption in the **original space** (that is, $y = \theta_0 + \theta_1x_1 + \theta_2x_2$) and the MSE criterion, estimate the parameters of the model $(\theta_0, \theta_1, \theta_2)$.

- (c) For each one of the 200 data points x_i of X , determine the associated estimate \hat{y}_i provided from the **model estimated in (b)** and compute the $MSE = \frac{1}{200} \sum_{i=1}^{200} (y_i - \hat{y}_i)^2$.
- (d) **Apply** the transformation $\varphi(x) = \begin{bmatrix} \varphi_1(x) \\ \varphi_2(x) \\ \varphi_3(x) \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \\ x_1 \cdot x_2 \end{bmatrix}$, on all x_i 's of X . Denoting by $x'_i (\in R^3)$ the image of x_i , form a new data set $X' = \{(y_i, x'_i), x'_i \in R^3, y_i \in R, i = 1, \dots, 200\}$.
- (e) **Adopting** the linear model assumption in the **transformed space** and the MSE criterion, estimate the parameters of the model.
- (f) For each one of the 200 data points x_i of X , determine the associated estimate \hat{y}_i provided from the **model estimated in (e)** and compute the $MSE = \frac{1}{200} \sum_{i=1}^{200} (y_i - \hat{y}_i)^2$.
- (g) Comment on the results obtained in (c) and (f).

Exercise 4 (python code + text):

- (a) **Generate** a set $X = \{(y_i, x_i), x_i \in R^2, y_i \in \{-1, +1\}, i=1, \dots, 2000\}$, as follows: Select 2000 points in the squared area $[-2, 2] \times [-2, 2]$ of the R^2 space, using the uniform distribution. All points that lie on the positive side of the curve $x_2^2 - x_1^2 = 0$, are assigned to the class "+1", while all the others are assigned to class "-1". Plot the data using different colors for points from different classes. In the sequel, pretend that you do not know how the data were generated. All you have at your disposal is the data set X .
- (b) **Apply** the transformation $\varphi(x) = \begin{bmatrix} \varphi_1(x) \\ \varphi_2(x) \end{bmatrix} = \begin{bmatrix} x_1^2 \\ x_2^2 \end{bmatrix}$, on all x_i 's of X . Denoting by x'_i the image of x_i , we form a new data set $X' = \{(y_i, x'_i), i=1, \dots, 2000\}$
- (c) **Plot** the x'_i 's using again different colors for points from different classes and compare the resulting plot with that of (a). Comment on them.
- (d) **Adopting** the linear model assumption in the transformed space and the MSE criterion, estimate the parameters of the model.

Exercise 5 (python code + text):

Consider a two-class classification problem (the classes y are denoted by +1 and -1), where the **first class** is modelled by a 2-dim. normal distribution with mean $m_1 = [0, 0]^T$

and covariance matrix $S_1 = \begin{bmatrix} 10 & 0 \\ 0 & 10 \end{bmatrix}$ and the second class is modelled by a 2-dim. normal distribution with mean $m_2 = [15, 15]^T$ and covariance matrix $S_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$

- Generate a data set consisting of 1000 data points that stem from class +1 and 20 data points that stem from class -1.
- Determine the separating line resulting from the LS criterion.
- Plot the data points as well as the separating line.
- Comment on the results obtained in (c).

Hints: For (a), use the relevant part of the python code presented in the class and has been uploaded in e-class. For (b), work as is shown in the example of the document titled “Detailed_derivation_of_LS_estimator” (e-class). For (c), use the relevant part of the python code presented in the class and has been uploaded in e-class. For (d), write down the MSE criterion taking into account that y ’s are in $\{+1, -1\}$. In addition, consider the estimates provided by the minimization of the LS criterion $y' = \theta^T \begin{bmatrix} 1 \\ x_i \end{bmatrix} =$

$[\theta_0 \ \theta_1 \ \theta_2] \begin{bmatrix} 1 \\ x_{i1} \\ x_{i2} \end{bmatrix}$, and compare them with the corresponding y ’s.

Exercise 6 (optional):

Verify the sum, the product and the Bayes rule for the discrete-valued case, using the relative frequency definition of the probability.