"Machine Learning and Computational Statistics"

4th Homework

Exercise 1:

Consider the **regression problem** y=g($\boldsymbol{x}$)+η

It is known that E[y|$\boldsymbol{x}$] is the minimum MSE estimate of y given $\boldsymbol{x}$. Consider the estimator $f(\boldsymbol{x};D)$.

(a) Under what conditions (theoretically) the quantity $\mathrm{E}_D[(f(\boldsymbol{x};D)-E[\mathrm{y}\,|\,\boldsymbol{x}])^2]$ becomes zero?

(b) Why this cannot be achieved in practice?

Exercise 2:

Consider a regression task modelled by the relation $y = g(\boldsymbol{x}) + \eta$. Let $f_{\boldsymbol{\theta}}$ be an estimator of g, parameterized by the vector $\boldsymbol{\theta}$. Let $Tr = \{(\boldsymbol{x}_n, y_n), n = 1, \ldots, N_1\}$ be the training set (the set that will be used for the estimation of $\boldsymbol{\theta}$**)** and $Te = \{(\boldsymbol{x}_n, y_n), n = 1, \ldots, N_2\}$ be the test (which will be used for testing the performance of $f_{\widehat{\boldsymbol{\theta}}}$ (where $\widehat{\boldsymbol{\theta}}$ is the estimate of $\boldsymbol{\theta}$ based on Tr)).

(a) What indicates a **large** error value on the training set Tr?

(b) What may indicate a **large** error value on the test set Te?

(c) What indicates a **small** error value on the training set Tr?

(d) What may indicate a **small** error value on the test set Te?

Exercise 3:

Consider a regression task $y = g(x) + \eta$, where y and x are modeled by the random variables y and x. The joint pdf of y and x is:

$p(x, y) = \frac{3}{2}$, for $x \in (0,1), y \in (x^2, 1)$.

Determine the optimum MSE estimate $E[y|x]$, for a given x, by performing the following steps:

(a) Verify that $p(x, y)$ is a pdf (prove that it integrates to 1).

(b) Compute the marginal pdf of x, $p_x(x)$.

(c) Compute the conditional pdf of y, given x.

(d) Compute and plot $E[y|x]$.

Hint: It is $\int_a^b x^n dx = [\frac{1}{n+1}x^{n+1}]_a^b = \frac{1}{n+1}b^{n+1} - \frac{1}{n+1}a^{n+1}$

Exercise 4 (python code + text):

Consider the **regression problem** (1-dep., 1-indep. variables)

$$y=g(x)+\eta$$

where y and x are jointly distributed according to the normal distribution $p(y,x) = N(\boldsymbol{\mu}, \Sigma)$

with $\boldsymbol{\mu} = \begin{bmatrix} \mu_y \\ \mu_x \end{bmatrix} = \begin{bmatrix} 2 \\ 2 \end{bmatrix}$ $\quad \Sigma = \begin{bmatrix} \sigma_y^2 & \sigma_{yx} \\ \sigma_{yx} & \sigma_x^2 \end{bmatrix} = \begin{bmatrix} 3 & 2 \\ 2 & 4 \end{bmatrix}$

(a) Determine $E[y|x]$ and plot the corresponding curve (recall the relevant theory concerning the normal distribution case).

(b) Generate 100 data sets D_i, i=1,…100, each one consisting of N=50 randomly selected pairs (y_n,x_n), $n=1,\ldots,N$, from $p(y,x)$.

(c) Adopt the linear estimator $f(x;D)$ and determine its instances $f(x;D_1)$,…, $f(x;D_{100})$, utilizing the LS criterion.

(d) Plot in a single figure **(i)** the lines corresponding to the above 100 estimates (blue color) and **(ii)** the line corresponding to the optimal MSE estimate (green color).

(e) Repeat steps (b)-(d) where now each data set consists of N=5000 points.

(f) Discuss the results (in your discussion, take into account the decomposition of the MSE to a variance and a bias term).

Exercise 5 (python code + text): Consider the setup of exercise 3. Generate a set D of $N = 100$ data pairs $\boldsymbol{z}_n=(y_n,x_n)$.

(a) For each x_n compute the optimal MSE estimate $\hat{y}_n$, $n = 1, \ldots, 100$ (use the results of exercise 3).

(b) Compute $\boldsymbol{\mu} = \begin{bmatrix} \mu_x \\ \mu_y \end{bmatrix} = \begin{bmatrix} \frac{1}{N}\sum_{n=1}^{N} x_n \\ \frac{1}{N}\sum_{n=1}^{N} y_n \end{bmatrix}$ and $\Sigma = \frac{1}{N}\sum_{n=1}^{N}(\boldsymbol{\mu} - \boldsymbol{z}_n)(\boldsymbol{\mu} - \boldsymbol{z}_n)^T$.

(c) Pretend that you do not know the true distribution that generates the data and you (**erroneously**) **assume** that the joint pdf of x and y, $p(x,y)$ is a normal one with mean and covariance matrix those computed in (b). Derive the optimum MSE estimate function for this assumption and compute the MSE estimate for each one of the 100 x_n's, $n = 1, \ldots, 100$.

(d) Discuss the results obtained from (a) and (c).

NOTE: Please give **brief explanations** in all **exercises**.