

MACHINE LEARNING-1 EXAM TRIAL QUESTION

1. LINEAR REGRESSION

Question 1. In this problem we will consider a regression problem. We will assume that we know certain aspects of the problem from the start and determine our modeling choices according to that knowledge.

Let's assume that we know the data generation process which is given by

$$y = x + \varepsilon,$$

where $y \in \mathbb{R}$, $x \in \mathbb{R}$, and $\varepsilon \sim \mathcal{N}(0, 1)$. According to this relation the training and test sets are obtained. For this particular problem we will work with the following sets.

$$Z_{\text{train}} = \{(-1, -1), (1, 1)\}$$

$$Z_{\text{test}} = \{(-2, -2)\}$$

We will try to model this underlying process with the linear model $M(x, w_1, w_2) = w_1 \times x + w_2$, where $w_1 \in \mathbb{R}$ and $w_2 \in \mathbb{R}$.

- What are the model parameters for the Ordinary Least Squares (OLS) solution?
- Assuming an uninformative prior $p(\mathbf{w}) \sim (\mathbf{0}, 10^6 \mathbf{I})$ on the weights $\mathbf{w} := [w_1, w_2]$, where $\mathbf{I} \in \mathbb{R}^{2 \times 2}$ is the identity matrix. What is the posterior $p(\mathbf{w}, Z_{\text{train}})$ for the weights of our model?
- What can you say about the Bayes and OLS solutions for the given problem.
- What is the expected error and variance of your error for both Bayes and OLS solutions for the test set, given that error measure is absolute error $(\hat{y} - y)$.

P.S.: You can work with one significant digit for your calculations.

Solution 1. For none of the problems matrix inversion is needed (due to identity inversion) every calculation can be done by hand.

- $\mathbf{w} = [1, 0]$. It is visible or can be computed by $(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y}$ where $\mathbf{X} := \begin{bmatrix} -1 & 1 \\ 1 & 1 \end{bmatrix}$ and $\mathbf{Y} := \begin{bmatrix} -1 \\ 1 \end{bmatrix}$.
- Noting $\alpha := 10^{-6}$ and $\beta = 1$ the posterior can be easily obtained by plugging in the formulation given in Bishop Eqns(3.53-3.54) $p(\mathbf{w}|Z_{\text{train}}) \sim \mathcal{N}(\mathbf{m}, \mathbf{V})$ where

$$\mathbf{m} := \beta \mathbf{V} \mathbf{X}^T \mathbf{Y} = \begin{bmatrix} -1 \\ 0 \end{bmatrix}$$

$$\mathbf{V} := (\alpha \mathbf{I} + \beta \mathbf{X}^T \mathbf{X})^{-1} = \begin{bmatrix} 0.5 & 0 \\ 0 & 0.5 \end{bmatrix}$$

- c With and uninformative prior given in the question posterior distribution reduces to the maximum likelihood solution and we know that the maximum likelihood solution is equal to OLS solution.
- d Expected Error for both of the models is 0 and the variance is 0 for OLS solution, but for the Bayes version we get the variance as 3.5 that is coming from the posterior predictive distribution.

$$p(y|x, Z_{\text{train}}, \alpha, \beta) \sim (\mathbf{m}\mathbf{X}^T, \sigma^2),$$

where $\mathbf{X} := \begin{bmatrix} -2 & 1 \end{bmatrix}$ and $\sigma = 1/\beta + \mathbf{X}\mathbf{V}\mathbf{X}^T = 3.5$