Machine Learning
Prof. S. Harmeling

Wednesday, 03. November 2021
DUE 23:55 Tuesday, 09. November 2021

Exercise set #4

Please submit your solutions in teams of two using the sciebo file-drop folder. The link is available in ILIAS. For the formatting please stick to the `submission_guideline.pdf` that you can find on sciebo. In the case of multiple uploads we will consider the latest. Uploads after the deadline will be deleted without further notice.

1. **Transformation of a variable: scaling and shifting**
   Assume x is a random variable with PDF $p_x(x)$. Define $y = ax + b$ being a scaled and shifted version of x for $a > 0$. Derive the PDF $p_y(y)$ using the transformation rule stated in Theorem 7.1 for the following two distributions:

   (a) $p_x(x) = \mathcal{U}(x|c, d)$, i.e., x is uniformly distributed on the interval $[c, d]$.

   (b) $p_x(x) = \mathcal{N}(x|\mu, \sigma^2)$, i.e., x is Gaussian distributed with mean μ and variance σ^2.

   25 points

2. **Bayesian linear regression**
   Assuming the model from the lecture

$$p(w) = \mathcal{N}(w|w_0, V_0)$$
$$p(y|X, w) = \mathcal{N}(y|Xw, \Sigma)$$

   derive the mean and variance of the posterior $p(w|X, y) = \mathcal{N}(w|w_n, V_n)$. For this write out the exponent of the product of prior and likelihood and bring it to the form of a multivariate second-order polynomial. Then apply Lemma 6.4 (3) from lecture 6 to get expressions for η and Λ. Reordering gives you expressions for w_n and V_n.

   30 points

3. **Boston housing data (programming task)**
   Goal of this exercise is to predict the price of the houses in Boston. Load the dataset from the text file `boston.txt`[1] using the function `np.genfromtxt`. Use the first 100 rows for testing, the next 50 rows for validation, i.e., for tuning hyperparameters, and the rest of the dataset for fitting your linear model. You do not have to introduce any additional features. For each dataset report the test mean squared error.

   (a) Which column is the target? Which columns are the features?

   (b) Fit a linear model for predicting the price using the MLE estimate w_{MLE}.

   (c) Fit a linear model for predicting the price using the ridge regression estimate w_{RIDGE}. Find a good choice for the regularization strength using the validation dataset.

   15 points

[1] Adapted from `http://lib.stat.cmu.edu/datasets/boston`

4. **Polynomial linear regression (programming task)**
Goal of this task is to fit a polynomial through data points $(x_1, y_1), \ldots, (x_n, y_n) \in \mathbb{R} \times \mathbb{R}$. Assume that the outcome $y = (y_1, \ldots, y_n)^T$ follows a normal distribution $\mathcal{N}(y|Xw, \sigma^2 I)$, where

$$X = \begin{bmatrix} | & | & | & | & | \\ 1 & x & x^2 & \ldots & x^d \\ | & | & | & | & | \end{bmatrix}$$

(a) Write a function that generates the matrix X for $x = (x_1, \ldots, x_n)^T$.

(b) Implement the estimator w_{MLE}.

(c) Implement a function that calculates the error $\text{MSE}(w) = \frac{1}{n}(Xw - y)^T(Xw - y)$.

(d) Try to find a good polynomial degree $d < 20$ that leads to a small validation error, i.e., the error on the validation dataset. Plot your best solution together with the training data and compute the error on the test dataset.

(e) Plot the training and test errors against the degree of the polynomial. A paper-pencil plot on squared paper is fine. What do you observe?

(f) Implement the estimator w_{RIDGE}.

(g) Find a good combination of d and λ that gives you a small validation error. Is the test error smaller than the test error of the optimal solution from (d)?

If you have problems with this exercise, let us know! Training, validation and test data are available in sciebo.

30 points