Dr. Peter Arndt
Dr. Konrad Völkel

Heinrich-Heine-Universität Düsseldorf
Winter term 2022/23

Machine Learning
Exercise Sheet 4
(3 Exercises, 90 Points)
Due: 8.11.2022, 10:00

Exercise 1: (30 Points)

Bayesian linear regression
Remember the model Bayesian linear regression model from the lecture, with multivariate normal distributions both for the measurement error and the prior:

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

In more detail: We assume a fixed covariance matrix Σ for the data distribution and only want to update the regression coefficients given in the vector w. The prior distribution for this is assumed to be a normal distribution with mean w_0 and covariance matrix V_0.

Derive the mean and variance of the posterior distribution $p(w|X, y)$.

[Lengthy hint: We know from the lecture that the posterior distribution is a normal distribution again; $p(w|X, y) = \mathcal{N}(w|w_n, V_n) = \exp(-\frac{1}{2}((w - w_n)^T V_n^{-1}(w - w_n))$ – the family of normal distributions is a conjugate family for normal distributions. Therefore it is enough to find the mean w_n and covariance matrix V_n of the posterior distribution.

For this write down the Bayesian fraction $p(w|X, y) \propto p(y|X, w)p(w|w_0, V_0)$ (disregard the denominator of the fraction). On the right hand side there is a product of two exponential functions. You can use the exponential laws and matrix operations to obtain an expression of the form $\exp(c + \eta^T w - \frac{1}{2} w^T \Lambda w)$ for some constant c, some vector η and some matrix Λ. You can disregard the constant, since it amounts to a factor that will be swallowed by the normalization constant in the end.

You then still need to bring this expression into the right form, finding w_n, V_n such that $\exp(\eta^T w - \frac{1}{2} w^T \Lambda w) = \exp(-\frac{1}{2}((w - w_n)^T V_n^{-1}(w - w_n))$. In your submission you can use without proof that $V_n = \Lambda^{-1}$ and $w_n = V_n \eta$ achieve this – but you are invited to check this for yourself!]

Exercise 2: (30 Points)

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.

1. Which column is the target? Which columns are the features?
2. Fit a linear model for predicting the price using the MLE estimate w_{MLE}.
3. 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.

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

Exercise 3: (30 Points)

Polynomial linear regression (programming task)

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

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

1. Write a function that generates the matrix X for $x = (x_1, \dots, x_n)^T$.
2. Implement the estimator w_{MLE}.
3. Implement a function that calculates the error $\text{MSE}(w) = \frac{1}{n}(Xw - y)^T(Xw - y)$.
4. 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.
5. 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?
6. Implement the estimator w_{RIDGE}.
7. 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.