Anna Krasilnikova 2562668
Mirza Misbah Mubeen Baig 2571567
Shahzain Mehboob 2571564

Exercise 5.1 – *Maximum Likelihood Estimation (MLE)*

a) Consider the density function of a univariate Gaussian distribution

$$p(x;\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right)$$

where μ is the mean and σ^2 is the variance.
Let's say you're given N samples (i.e. $x_1, x_2, , x_3, \dots, x_N$) which are drawn from the above stated distribution. Also, you can assume that these samples are i.i.d (i.e. independent and identically distributed). Derive the MLE step-by-step for:

i) mean (μ)

We know, for Maximum likelihood for some $N(\mu,\sigma^2)$ given $\{x\}_i$, for $i \in 1..n$ we can make following observations:

$$L(\mu,\sigma^2,\{x_i\}) = \prod N(\mu,\sigma^2) = \prod \frac{1}{\sqrt{2\pi\sigma^2}} e^{\left(-\frac{1}{2\sigma^2}(x_i-\mu)^2\right)} = (2\pi\sigma^2)^{-\frac{n}{2}} e^{-\frac{1}{2\sigma^2}\Sigma(x_i-\mu)^2}$$

Overall, we can represent the log as:

$$\log\big(L(\mu,\sigma^2,\{x_i\})\big) = \log(2\pi\sigma^2)^{\frac{n}{2}} + \log\left(e^{-\frac{1}{2\sigma^2}\Sigma(x_i-\mu)^2}\right) = -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum(x_i-\mu)^2$$

Now, having those transformations, we can proceed to the estimate for μ:

$$\big[\log\big(L(\mu,\sigma,\{x_i\})\big)\big]'_\mu = 0$$

$$\left(-\frac{\pi}{2}\log(2\pi\sigma^2)\right)'_\mu - \left(\frac{1}{2\sigma^2}\sum_i^n (x_i-\mu)^2\right)'_\mu = 0$$

The first part is clearly a zero, thus we can transform the equation into the following form:

$$-\frac{1}{2\sigma^2}\sum_i^n(-2x_i+2\mu) = 0$$

Now, dividing by two, we can get:

$$\frac{1}{\sigma^2}\sum_i^n(\mu - x_i) = 0$$

Hence,

$$\sum_i^n x_i - \sum_i^n \mu = 0, \qquad \Longrightarrow \sum_i^n x_i = \mu n \Longrightarrow \mu = \frac{\sum_i^n x_i}{n}$$

ii) variance (σ^2)

First, we can show following:

$$\left(-\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_i^n(x_i-\mu)^2\right)'_{\sigma^2} = 0$$

Perform some transformations:

$$= -\frac{n}{2}[\log(2\pi) + \log(\sigma^2)]'_{\sigma^2} = -\frac{n}{2\sigma^2}$$

$$\left(-\frac{1}{2\sigma^2}\sum_i^n(x_i-\mu)^2\right)'_{\sigma^2} = -\frac{1}{2}\sum_i^n(x_i-\mu)^2\,(-\frac{1}{(\sigma^2)^2}$$

Then we can show following:

Anna Krasilnikova 2562668
Mirza Misbah Mubeen Baig 2571567
Shahzain Mehboob 2571564

$$-\frac{n}{2\sigma^2}+\frac{1}{2}\sum_{i}^{n}(x_i-\mu)^2\frac{1}{\sigma^4}=0$$

$$\Rightarrow \frac{1}{2\sigma^2}\left(\frac{\sum_i^n(x_i-\mu)^2}{\sigma^2}\right)=0$$

Given that $\frac{\sum_i^n(x_i-\mu)^2}{\sigma^2}=n$, we can get σ^2 out of this equation:

$$\sigma^2=\frac{\sum_i^n(x_i-\mu)^2}{n}$$

b) *Consider the linear regression problem from the lecture. The ground truth y_i can be considered as being generated by $y_i=w_{gt}^T x_i+\epsilon_i$, where ϵ_i is a random noise which follows a standard Gaussian distribution and w_{gt} denotes the ground truth weight. Prove that MLE of w with $p(y_i|x_i)$ is equivalent to minimizing the MSE for w from the lecture.*

We know that y_i is presented as $y_i=w_{gt}^T x_i+\epsilon_i$.

Thus, we can present $p(y_i|x_i)$ as following:

$$p(y_i|x_i)=\frac{1}{\left(\sqrt{2\pi\sigma^2}\right)^n}e^{-\frac{1}{2\sigma^2}\sum_i^n(y_i-wx_i)^2}=\frac{1}{\left(\sqrt{2\pi\sigma^2}\right)}e^{-\frac{1}{2\sigma^2}\|Y-WX\|_2^2}$$

Now we can apply the log:

$$-\frac{1}{2}\log 2\pi\sigma^2-\frac{1}{\sigma^2}\|Y-WX\|_2^2$$

Overall, maximizing the $\log(L(x,\mu,\sigma^2))$ will be equivalent to minimizing $\|Y-WX\|_2^2$, which is nothing, but MSE problem. This concludes the proof.

Exercise 5.2 – *Validation and Cross-Validation*

a) *Why would one need cross-validation instead of the holdout method?*

Holdout technique divides the data in to two sets, the training set and the testing sets. The advantage of holdout technique is less computation time due to residual methods and disadvantage for holdout technique is variance of evaluation due to data division because data points which ends up in the training set also end up in the test set.

K-fold cross validation is a good replacement of holdout method, because in k-fold technique we divide the data in to K sets and the same holdout technique is repeated K times. K-fold cross validation works in such a way that in each iteration all other sets that are not used in testing are put up together to create a training set which ends up in having one test set and k-1 training test and every data points becomes a part of test set only once and training sets k-1 times. In the end the average of errors is computed from all trails. We can reduce the variance by increasing number of k but it will induce computation time as for every trial the same algorithm executed from the beginning.

Scenarios for K-fold cross validation are, we can use cross validation when we need to assess the predictive performance of the models and to judge how they perform outside the sample to a new data set or when we have to check that how well a model generalizes to new data or to estimate any quantitative measure of fit that is appropriate for both data and model.

Example of K-fold cross validation is, imagine we have a data set with 6 data points and we selected a value k=3 for the number of folds used to split our data. This means we will shuffle our data and split them into 3 equally big groups of 2 data points namely Fold1, Fold2, Fold3. Each fold given the chance to be the held out testing set.

Model1: Trained on (Fold1 + Fold2), Tested on Fold3

Model2: Trained on (Fold2 + Fold3), Tested on Fold1

Anna Krasilnikova 2562668
Mirza Misbah Mubeen Baig 2571567
Shahzain Mehboob 2571564

Model3: Trained on (Fold1 + Fold3), Tested on Fold2.

b) Assume that we want to find an optimal capacity of our model for the task of linear regression, with possible choices for the order of the polynomial as: {1, 5, 9}.
Now, assume that we want to do this hyperparameter selection using k-fold crossvalidation (with k = 5), instead of the holdout method. Given this setting, pictorially explain the steps that are involved in this 5-fold cross-validation, along with a brief explanation. Also, explain how would you compute a single final score (e.g. MSE) for each of the hyperparameters (i.e. order of the polynomial), so that we can compare the performance of these models and choose the best one.

For the order of the polynomial 1 and given K = 5:

Step # 1

TEST	TRAIN	TRAIN	TRAIN	TRAIN

In this step we will select the **first** portion of the training data as a validation set and we will use rest of the training data for training the model. For polynomial of degree 1, we will train our model as we train for a simple regression model and then calculate the MSE on the validation set that we set aside.

$$Y = mx + b \ (for\ polynomial\ of\ degree\ 1)$$
$$E^1 = \frac{1}{n}\sum_{i=1^n}\left(y_i^{test} - y_i^{train}\right)^2$$

Step # 2

TRAIN	TEST	TRAIN	TRAIN	TRAIN

In this step we will select the **second** portion of the training data as a validation set and we will use rest of the training data for training the model. For polynomial of degree 1, we will train our model as we train for a simple regression model and then calculate the MSE on the validation set that we set aside.

$$Y = mx + b \ (for\ polynomial\ of\ degree\ 1)$$
$$E^2 = \frac{1}{n}\sum_{i=1^n}\left(y_i^{test} - y_i^{train}\right)^2$$

Step # 3

TRAIN	TRAIN	TEST	TRAIN	TRAIN

In this step we will select the **third** portion of the training data as a validation set and we will use rest of the training data for training the model. For polynomial of degree 1, we will train our model as we train for a simple regression model and then calculate the MSE on the validation set that we set aside.

$$Y = mx + b \ (for\ polynomial\ of\ degree\ 1)$$
$$E^3 = \frac{1}{n}\sum_{i=1^n}\left(y_i^{test} - y_i^{train}\right)^2$$

Step # 4

TRAIN	TRAIN	TRAIN	TEST	TRAIN

In this step we will select the **fourth** portion of the training data as a validation set and we will use rest of the training data for training the model. For polynomial of degree 1, we will train our model as we train

Anna Krasilnikova 2562668
Mirza Misbah Mubeen Baig 2571567
Shahzain Mehboob 2571564

for a simple regression model and then calculate the MSE on the validation set that we set aside.

$$Y = mx + b\ (for\ polynomial\ of\ degree\ 1)$$
$$E^4 = \frac{1}{n}\sum_{i=1^n}\left(y_i^{test} - y_i^{train}\right)^2$$

Step # 5

TRAIN	TRAIN	TRAIN	TRAIN	TEST

In this step we will select the **fifth** portion of the training data as a validation set and we will use rest of the training data for training the model. For polynomial of degree 1, we will train our model as we train for a simple regression model and then calculate the MSE on the validation set that we set aside.

$$Y = mx + b\ (for\ polynomial\ of\ degree\ 1)$$
$$E^5 = \frac{1}{n}\sum_{i=1^n}\left(y_i^{test} - y_i^{train}\right)^2$$

Now we will compute the average over all the Errors we have computed by setting the hyperparameter of polynomial of degree 1:

$$Error^{total} = \frac{1}{k}\sum_{k=1^k} E^i$$

For the order of the polynomial 5 and given K = 5:

Step # 1

TEST	TRAIN	TRAIN	TRAIN	TRAIN

In this step we will select the **first** portion of the training data as a validation set and we will use rest of the training data for training the model. Here we will transform our input variables into polynomial degree of 5 and train our model as we train for a linear regression model and then calculate the MSE on the validation set that we set aside.

$$Y = m_1x_1 + m_2x_2 + m_3x_3 + m_4x_4 + m_5x_5 + b\ (for\ polynomial\ of\ degree\ 5)$$
$$Where\ x_1 = x, x_2 = x^2, \quad x_3 = x^3, x_4 = x^4, and\ x_5 = x^5$$
$$E^1 = \frac{1}{n}\sum_{i=1^n}\left(y_i^{test} - y_i^{train}\right)^2$$

Step # 2

TRAIN	TEST	TRAIN	TRAIN	TRAIN

In this step we will select the **second** portion of the training data as a validation set and we will use rest of the training data for training the model. Here we will transform our input variables into polynomial degree of 5 and train our model as we train for a linear regression model and then calculate the MSE on the validation set that we set aside.

$$Y = m_1x_1 + m_2x_2 + m_3x_3 + m_4x_4 + m_5x_5 + b\ (for\ polynomial\ of\ degree\ 5)$$
$$Where\ x_1 = x, x_2 = x^2, \quad x_3 = x^3, x_4 = x^4, and\ x_5 = x^5$$

Anna Krasilnikova 2562668
Mirza Misbah Mubeen Baig 2571567
Shahzain Mehboob 2571564

$$E^2 = \frac{1}{n}\sum_{i=1^n}\left(y_i^{test} - y_i^{train}\right)^2$$

Step # 3

TRAIN	TRAIN	TEST	TRAIN	TRAIN

In this step we will select the **third** portion of the training data as a validation set and we will use rest of the training data for training the model. Here we will transform our input variables into polynomial degree of 5 and train our model as we train for a linear regression model and then calculate the MSE on the validation set that we set aside.

$$Y = m_1x_1 + m_2x_2 + m_3x_3 + m_4x_4 + m_5x_5 + b\ (for\ polynomial\ of\ degree\ 5)$$
$$Where\ x_1 = x, x_2 = x^2, \quad x_3 = x^3, x_4 = x^4, and\ x_5 = x^5$$
$$E^3 = \frac{1}{n}\sum_{i=1^n}\left(y_i^{test} - y_i^{train}\right)^2$$

Step # 4

TRAIN	TRAIN	TRAIN	TEST	TRAIN

In this step we will select the **fourth** portion of the training data as a validation set and we will use rest of the training data for training the model. Here we will transform our input variables into polynomial degree of 5 and train our model as we train for a linear regression model and then calculate the MSE on the validation set that we set aside.

$$Y = m_1x_1 + m_2x_2 + m_3x_3 + m_4x_4 + m_5x_5 + b\ (for\ polynomial\ of\ degree\ 5)$$
$$Where\ x_1 = x, x_2 = x^2, \quad x_3 = x^3, x_4 = x^4, and\ x_5 = x^5$$
$$E^4 = \frac{1}{n}\sum_{i=1^n}\left(y_i^{test} - y_i^{train}\right)^2$$

Step # 5

TRAIN	TRAIN	TRAIN	TRAIN	TEST

In this step we will select the **fifth** portion of the training data as a validation set and we will use rest of the training data for training the model. Here we will transform our input variables into polynomial degree of 5 and train our model as we train for a linear regression model and then calculate the MSE on the validation set that we set aside.

$$Y = m_1x_1 + m_2x_2 + m_3x_3 + m_4x_4 + m_5x_5 + b\ (for\ polynomial\ of\ degree\ 5)$$
$$Where\ x_1 = x, x_2 = x^2, \quad x_3 = x^3, x_4 = x^4, and\ x_5 = x^5$$
$$E^5 = \frac{1}{n}\sum_{i=1^n}\left(y_i^{test} - y_i^{train}\right)^2$$

Now again we will compute the average over all the Errors we have computed by setting the hyperparameter of polynomial of degree 5:

$$Error^{total} = \frac{1}{k}\sum_{k=1^k} E^i$$

For the order of the polynomial 9 and given K = 5:

Anna Krasilnikova 2562668
Mirza Misbah Mubeen Baig 2571567
Shahzain Mehboob 2571564

Step # 1

TEST	TRAIN	TRAIN	TRAIN	TRAIN

In this step we will select the **first** portion of the training data as a validation set and we will use rest of the training data for training the model. Here also we will transform our input variables into polynomial degree of 9 and train our model as we train for a linear regression model and then calculate the MSE on the validation set that we set aside.

$$Y = m_1x_1 + m_2x_2 + m_3x_3 + m_4x_4 + m_5x_5 + m_6x_6 + m_7x_7 + m_8x_8 + m_9x_9 + b\ (for\ polynomial\ of\ degree\ 9)$$

$$Where\ x_1 = x, x_2 = x^2,$$

$$x_3 = x^3, x_4 = x^4, x_5 = x^5, x_6 = x^6, x_7 = x^7, x_8 = x^8\ and\ x^9 = x^9$$

$$E^1 = \frac{1}{n}\sum_{i=1^n}\left(y_i^{test} - y_i^{train}\right)^2$$

Step # 2

TRAIN	TEST	TRAIN	TRAIN	TRAIN

In this step we will select the **second** portion of the training data as a validation set and we will use rest of the training data for training the model. Here also we will transform our input variables into polynomial degree of 9 and train our model as we train for a linear regression model and then calculate the MSE on the validation set that we set aside.

$$Y = m_1x_1 + m_2x_2 + m_3x_3 + m_4x_4 + m_5x_5 + m_6x_6 + m_7x_7 + m_8x_8 + m_9x_9 + b\ (for\ polynomial\ of\ degree\ 9)$$

$$Where\ x_1 = x, x_2 = x^2,$$

$$x_3 = x^3, x_4 = x^4, x_5 = x^5, x_6 = x^6, x_7 = x^7, x_8 = x^8\ and\ x^9 = x^9$$

$$E^2 = \frac{1}{n}\sum_{i=1^n}\left(y_i^{test} - y_i^{train}\right)^2$$

Step # 3

TRAIN	TRAIN	TEST	TRAIN	TRAIN

In this step we will select the **third** portion of the training data as a validation set and we will use rest of the training data for training the model. Here also we will transform our input variables into polynomial degree of 9 and train our model as we train for a linear regression model and then calculate the MSE on the validation set that we set aside.

$$Y = m_1x_1 + m_2x_2 + m_3x_3 + m_4x_4 + m_5x_5 + m_6x_6 + m_7x_7 + m_8x_8 + m_9x_9 + b\ (for\ polynomial\ of\ degree\ 9)$$

$$Where\ x_1 = x, x_2 = x^2,$$

$$x_3 = x^3, x_4 = x^4, x_5 = x^5, x_6 = x^6, x_7 = x^7, x_8 = x^8\ and\ x^9 = x^9$$

$$E^3 = \frac{1}{n}\sum_{i=1^n}\left(y_i^{test} - y_i^{train}\right)^2$$

Step # 4

Anna Krasilnikova 2562668
Mirza Misbah Mubeen Baig 2571567
Shahzain Mehboob 2571564

TRAIN	TRAIN	TRAIN	TEST	TRAIN

In this step we will select the **fourth** portion of the training data as a validation set and we will use rest of the training data for training the model. Here also we will transform our input variables into polynomial degree of 9 and train our model as we train for a linear regression model and then calculate the MSE on the validation set that we set aside.

$$Y = m_1x_1 + m_2x_2 + m_3x_3 + m_4x_4 + m_5x_5 + m_6x_6 + m_7x_7 + m_8x_8 + m_9x_9 + b\ (for\ polynomial\ of\ degree\ 9)$$

$$Where\ x_1 = x, x_2 = x^2,\ x_3 = x^3, x_4 = x^4, x_5 = x^5, x_6 = x^6, x_7 = x^7, x_8 = x^8\ and\ x^9 = x^9$$

$$E^4 = \frac{1}{n}\sum_{i=1^n}\left(y_i^{test} - y_i^{train}\right)^2$$

Step # 5

TRAIN	TRAIN	TRAIN	TRAIN	TEST

In this step we will select the **fifth** portion of the training data as a validation set and we will use rest of the training data for training the model. Here also we will transform our input variables into polynomial degree of 9 and train our model as we train for a linear regression model and then calculate the MSE on the validation set that we set aside.

$$Y = m_1x_1 + m_2x_2 + m_3x_3 + m_4x_4 + m_5x_5 + m_6x_6 + m_7x_7 + m_8x_8 + m_9x_9 + b\ (for\ polynomial\ of\ degree\ 9)$$

$$Where\ x_1 = x, x_2 = x^2,\ x_3 = x^3, x_4 = x^4, x_5 = x^5, x_6 = x^6, x_7 = x^7, x_8 = x^8\ and\ x^9 = x^9$$

$$E^5 = \frac{1}{n}\sum_{i=1^n}\left(y_i^{test} - y_i^{train}\right)^2$$

Now again we will compute the average over all the Errors we have computed by setting the hyperparameter of polynomial of degree 5:

$$Error^{total} = \frac{1}{k}\sum_{k=1^k} E_i$$

In the end we have three Total Error terms that is calculated from three different hyperparameters of 1, 3 and 9 order of polynomial with k-fold cross validation (K=5) therefore we can now select one of the those values which gives the lowest error and then apply that model on our real test data.

Exercise 5.3 – *Logistic Regression*

Why would one need or choose (multinomial) logistic regression over linear regression irrespective of the usage of regularization? Explain.

Linear regression needs to determine linearity between dependent and independent variable whereas this is not the case with logistic regression. It also demands all the dependent variable to be a numeric value that means either categorized or grouped (continuous) and on the other hand independent variable has to be correlated with each other. In contrast with the logistic regression, the variable must not have needed to be correlated with each other. Lastly, Linear regression is based on least square estimation which minimizes the sum of the squared distances of each observed response to its fitted value, so

Anna Krasilnikova 2562668
Mirza Misbah Mubeen Baig 2571567
Shahzain Mehboob 2571564

Linear regression will be used when your response variable is continuous. For instance, age, time, weight, height, number of hours, etc. While logistic regression is based on Maximum Likelihood Estimation which maximizes the Probability of variable Y given X (likelihood). So Logistic regression will be used when the response variable is categorical in nature e.g. Days, Boolean, gender, ranks, cities, colors and all other similar type.