Polynomial regression

1. Two models are proposed to fit data:

   Model 1: $y = w_0 + w_1 x$

   Model 2: $y = w_0 + w_1 x + w_2 x^2 + \cdots + w_9 x^9$

   (a) **Model 1**, when compared to **Model 2**, for the same train dataset of about 30 entries is expected to have:

   A. Lower MSE on training dataset
   B. Higher MSE on training dataset
   C. Lower MSE on test dataset
   D. Higher MSE on test dataset
   E. The difference for training dataset cannot be determined.
   F. The difference for test dataset cannot be determined.

   Solution: Model 2 is expected to overfit on train dataset.

   (b) Adding 100 entries to the train dataset beyond the initial entries, for **Model 1**:

   ○ Improves significantly MSE on test dataset
   ○ Improves significantly MSE on train dataset
   ○ Worsen significantly MSE on test dataset
   ○ Worsen significantly MSE on train dataset
   √ **Change in MSE on test dataset is negligible**
   √ **Change in MSE on train dataset is negligible**

   Solution: Model 1 is either fit or underfit and cannot improve significantly.

   (c) Adding 100 entries to train dataset beyond the initial entries, for **Model 2**:

   √ **Improves MSE on test dataset**
   ○ Improves MSE on train dataset
   ○ Worsen MSE on test dataset
   √ **Worsen MSE on train dataset**
   ○ Change in MSE on test dataset is negligible
   ○ Change in MSE on train dataset is negligible

   Solution: Model 2 is overfit that reduces with increase in the size of the dataset.

   (d) The number of entries in train dataset is 10. In this case:

   A. Model 1 is likely to overfit the data
   B. Model 2 is likely to overfit the data
   C. Model 1 is likely to underfit the data
   D. Model 2 is likely to underfit the data

   Solution: Model 1 may either underfit or fit the data (2 uncertain options). Model 2 for sure overfits the data.

(e) **Model 1** is trained on 100 dataset entries and has similar train and test MSE

√ **The model is in its optimal performance.**

○ Regularization is recommended to improve performance.

○ Adding dataset entries is recommended to improve performance.

○ Neither answer is correct.

Solution: Only change of the model can improve the performance.

(f) **Model 2** is trained on 100 dataset entries and train MSE is significantly better then test MSE.

√ **The model is likely to overfit**

○ The model is likely to underfit

○ The model is in its optimal performance.

√ **Regularization is recommended to improve performance.**

√ **Adding dataset entries is recommended to improve performance.**

(g) **Model 3:** $y = w_0 + w_1 x + w_2 x^2 + \cdots + w_N x^N$ is proposed

○ N is model parameter

√ **N is model hyper-parameter**

○ For the selection of optimal N, the N with the minimal MSE performance on the train dataset is selected.

○ For the selection of optimal N, the N with the minimal MSE performance on the test dataset is selected.

√ **For the selection of optimal N, the N with the minimal MSE performance on the validation dataset is selected.**

Regularization

2. During the training of the model, the following weights were learned:

	A	**B**
w_0	-0.93232	-0.93232
w_1	-0.39239	0.641779
w_2	-0.24492	-0.319227
w_3	-0.48193	1.043743
w_4	-0.24317	-0.340061
w_5	0.061323	-0.78238

A. Column A is for L_2-regularized weights, and column B is without regularization.

B. Column **A** is without regularization and column **B** is for L_2-regularized weights.

C. Columns cannot be clearly attributed if regularized or not.

Solution: Note the significant difference in $\sum_i w_i^2$.

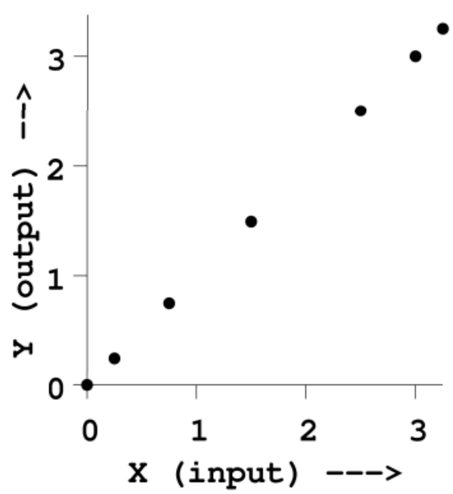
3. λ is termed as regularization parameter. For sufficiently high λ:
   A. The model tend to underfit
   B. The model tend to overfit
   C. The model tend to reach its optimal performance
   D. The model is not changed

Solution: The factor $\lambda \sum_i w_i^2$ dominates the original loss function.

Linear Regression

Linear regression model is $y = w_0 + w_1 x$.

4. Consider the following data



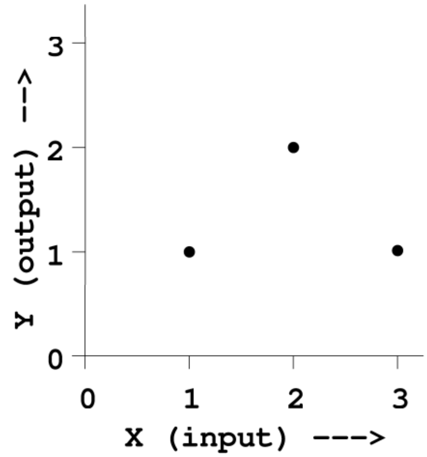
(a) What is the expected MSE on this data?

Solution: Without calculation, $MSE = 0$ since it is a linear line.

(b) 3 points on the right are test set and 3 points on the left are train test. What is MSE of the both?

Solution: Without calculation, $MSE = 0$ since it is a linear line.

5. Consider the following data:



X	Y
1	1
2	2
3	1

(a) What is the expected MSE on this data? (Hint: by symmetry it is clear that the best fit to the three datapoints is a horizontal line).

Solution: Following the hint, $w_1 = 0$ and $w_0 = ?$

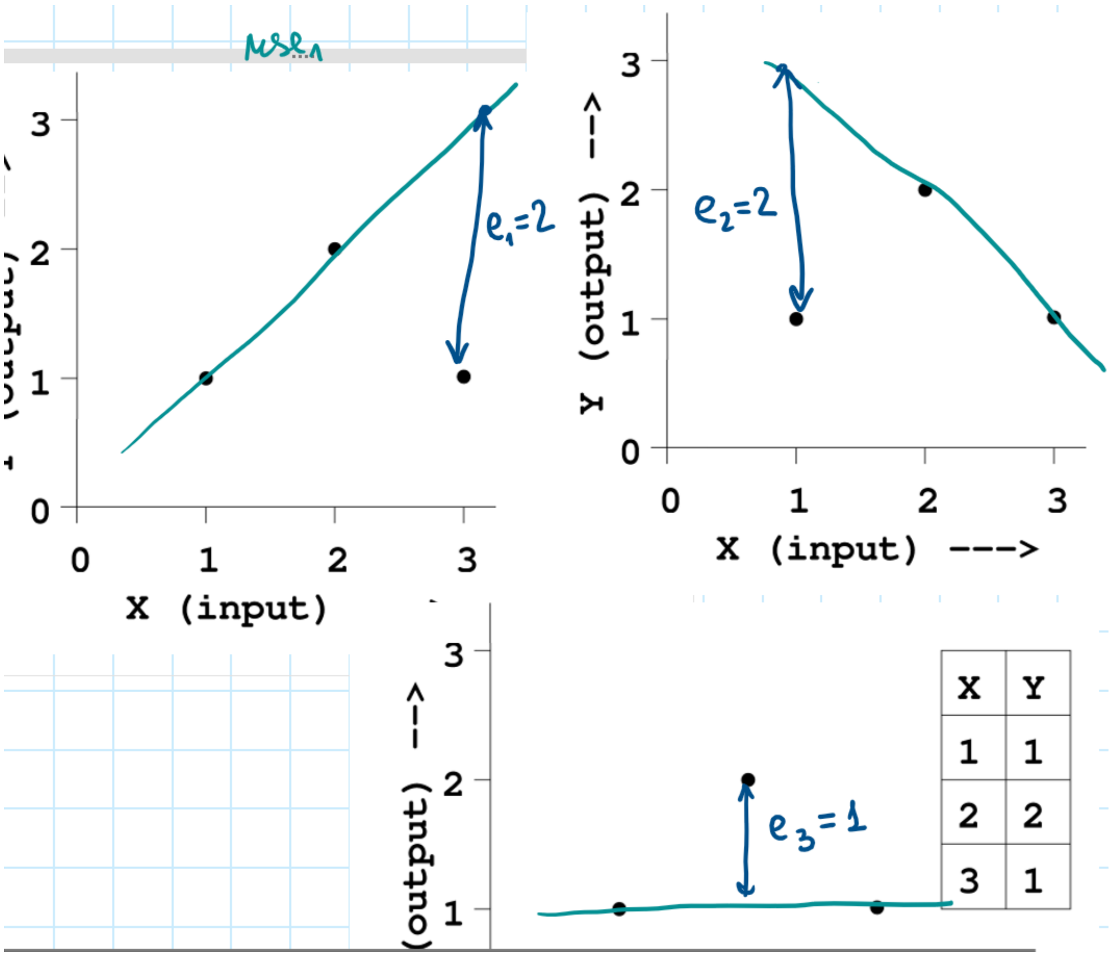
$$\begin{aligned}
MSE &= e_1^2 + e_2^2 + e_3^2 \\
&= (w_0 - 1)^2 + (w_0 - 2)^2 + (w_0 - 1)^2 \\
&= w_0^2 + w_0^2 + w_0^2 - 2w_0 - 4w_0 - 2w_0 + 1 + 4 + 1 \\
&= 3w_0^2 - 8w_0 + 6 \\
MSE_2 &= \frac{1}{3}\left(e_1^2 + e_2^2 + e_3^2\right) = \frac{1}{3}MSE \\
\frac{d}{dw_0}MSE &= 6w_0 - 8 = 0 \\
w_0 &= \frac{4}{3} = 1\frac{1}{3} \\
\Rightarrow MSE(w_0) &= 3\left(\frac{4}{3}\right)^2 - 8\frac{4}{3} + 6 = \frac{2}{3}
\end{aligned}$$

(b) (9 points) What is the MSE of leave-one-out cross-validation (3-fold in this case) of running linear regression on this data?

Solution: Each time one point is for test and two others are for train.

By basic analytical geometry,

$$
\begin{aligned}
MSE_1 &= 2^2 = 4 \\
MSE_2 &= 2^2 = 4 \\
MSE_3 &= 1 \\
\overline{MSE} &= \frac{1}{3}(4 + 1 + 4) = 3
\end{aligned}
$$



Logistic Regression

6. Logistic regression model is given by

$$\hat{y} = \frac{1}{1 + \exp(-\mathbf{x}^T \mathbf{w})}$$

For weights $w_0 = -15, w_1 = 3, w_2 = 5$, what is the prediction probability for $\hat{y} = 1$ for $(x_1, x_2) = (1, 2)$?

Solution:

$$\hat{y} = \frac{1}{1 + \exp(-(-15 + 3 \cdot 1 + 5 \cdot 2))} = 0.11920292202211755$$