Aprendizagem 2021

Lab 4: Linear Regression

Practical exercises

1. Consider the following training data:

	y_1	y_2	output
x_1	1	1	1.4
x_2	2	1	0.5
x_3	1	3	2
x_4	3	3	2.5

a) Find the closed form solution for a linear regression minimizing the sum of squared errors

b) Predict the target value for $x_{new} = [2\ 3]^T$

c) Sketch the predicted three-dimensional hyperplane

d) Compute the MSE and MAE produced by the linear regression

e) Are there biases on the residuals against y1? And y2?

f) Compute the closed form solution considering Ridge regularization term with $\lambda = 0.2$.

g) Compare the hyperplanes obtained using ordinary least squares and Ridge regression.

h) Why is Lasso regression suggested for data spaces of higher dimensionality?

2. Consider the following training data where *output* is an ordinal variable

	y_1	y_2	output
x_1	1	1	1
x_2	2	1	1
x_3	1	3	0
x_4	3	3	0

a) Find a linear regression using the closed form solution

b) Assuming the output threshold θ=0.5, use the regression to classify $x_{\text{new}} = [2\ 2.5]^T$

3. [*optional*] Consider the following data to learn a model $z = w_1 y_1 + w_2 y_2 + \varepsilon$, where $\varepsilon \sim N(0, 0.1)$

	y_1	y_2	output
x_1	3	-1	2
x_2	4	2	1
x_3	2	2	1

Compare:

a) $\boldsymbol{w} = [w_1\ w_2]^T$ using the maximum likelihood approach

b) $\boldsymbol{w}$ using the Bayesian approach, assuming $p(\boldsymbol{w}) = N\left(\boldsymbol{w} \mid \boldsymbol{u} = [0\ 0], \boldsymbol{\sigma} = \begin{bmatrix} 0.2 & 0 \\ 0 & 0.2 \end{bmatrix}\right)$

4. Identify a transformation to aid the linearly modelling of the following data points.

Sketch the predicted surface.

	y_1	y_2	output
x_1	-0.95	0.62	0
x_2	0.63	0.31	0
x_3	-0.12	-0.21	1
x_4	-0.24	-0.5	0
x_5	0.07	-0.42	1
x_6	0.03	0.91	0
x_7	0.05	0.09	1
x_8	-0.83	0.22	0

5. Consider logarithmic and quadratic transformations:
$\varphi_1(x_1) = log(x_1), \quad \varphi_2(x_1) = {x_1}^2$

a) Plot both of the closed form regressions.

b) Which one minimizes the sum of squared errors on the original training data

	input	*output*
x_1	3	1.5
x_2	4	9.3
x_3	6	23.4
x_4	10	45.8
x_5	12	60.1

6. Select the criteria that promotes a *smoother* regression model:

- Applying Lasso and Ridge regularization to linear regression models
- Increasing the depth of a decision tree regressor
- Increasing the *k* of a *k*NN regressor
- Parameterizing a *k*NN regressor with uniform weights instead of distance-based weights

Programming quests

7. Consider the *housing* dataset available at https://web.ist.utl.pt/~rmch/dscience/data/housing.arff

a) Compare the determination coefficient of the non-regularized, Lasso and Ridge linear regression on the housing data
Resource: https://www.pluralsight.com/guides/linear-lasso-ridge-regression-scikit-learn

b) Compare the MAE and RMSE of linear, kNN and decision tree regressors on housing

8. Learn and interpret the logistic regression model for the Iris dataset
Resource: https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LogisticRegression.html