Foundations of Machine Learning - Exercise (SS 25)

Assignment 4: Ridge Regression, Correlation, and Classification Metrics

Arvindh Arunbabu
arvindh.arun@ki.uni-stuttgart.de

Nadeen Fathallah
Nadeen.Fathallah@ki.uni-stuttgart.de

Akram Sadat Hosseini
Akram.Hosseini@ki.uni-stuttgart.de

Farane Jalali
farane.jalali-farahani@ki.uni-stuttgart.de

Osama Mohammed
osama.mohammed@ki.uni-stuttgart.de

Jiaxin Pan
jiaxin.pan@ki.uni-stuttgart.de

Tim Schneider
tim.schneider@ki.uni-stuttgart.de

Jingcheng Wu
jingcheng.wu@ki.uni-stuttgart.de

Daniel Frank
daniel.frank@ki.uni-stuttgart.de

Submit your theoretical solution in ILIAS as a single PDF file.[1] Make sure to list the full names of all participants, matriculation number, study program, and B.Sc. or M.Sc on the first page. Optionally, you can *additionally* upload source files (e.g., PPTX files). Submit your programming task in ILIAS as a single Jupyter notebook. If you have any questions, feel free to ask them in the exercise forum in ILIAS.

Submission is open until Monday, 19th of May, 12:00 noon.

[1] Your drawing software probably allows exporting as PDF. An alternative option is to use a PDF printer. If you create multiple PDF files, use a merging tool (like pdfarranger) to combine the PDFs into a single file.

Task 1: Ridge Regression

In the lecture it was derived that linear regression minimizes the *mean squared error (MSE)* loss with the minimizer $\mathbf{w}^* = (X^T X)^{-1} X^T \mathbf{y}$. The idea of *ridge regression* is to augment the loss with an additional regularization term, the euclidean norm $\|\mathbf{w}\|_2^2$ of the coefficients.

1. **Task** : Derive a closed-form solution to the loss $\mathcal{L} = \|X\mathbf{w} - \mathbf{y}\|_2^2 + \lambda\|\mathbf{w}\|_2^2$.
2. **Task** : The solution to a linear system $Ax = b$, with symmetric, positive definite $A \in \mathbb{R}^{m\times m}$ be solved in $\mathcal{O}(n^3)$ with a Cholesky decomposition. What is the computational complexity of training a linear regression with n datapoints of dimensionailty d?
3. **Task** : If $A \in \mathbb{R}^{d\times d}$, $U \in \mathbb{R}^{d\times n}$, $C \in \mathbb{R}^{n\times n}$, $V \in \mathbb{R}^{n\times d}$, and A, C are invertible, then

$$(A + UCV)^{-1} = A^{-1} - A^{-1}U\left(C^{-1} + VA^{-1}U\right)^{-1} VA^{-1},$$

   which known as the *Woodbury Matrix indentity*. Apply it to the solution of ridge regression.
4. **Task** Again, think about the computational complexity.
5. **Task** : What does a single entry $(XX^T)_{ij}$ of the matrix represent?

Task 2: Difference between Correlation and Covariance

Please see the attached Jupyter notebook.

Task 3: Classification Metrics

F-Score

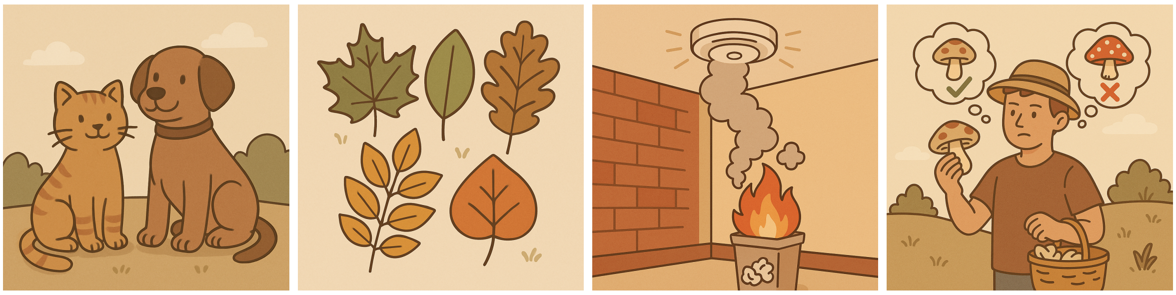
1. **Task** What's the *minimum* and *maximum* possible F_1-score? Shortly describe a classifier that achieves that score in a binary classification task.
2. **Task** In the lecture we defined $F_1 = \frac{2 \cdot \text{Prec} \cdot \text{Rec}}{\text{Prec} + \text{Rec}}$. What's the F_1 in case of $\text{Prec} = \text{Rec} = 0$? *(Hint: $\frac{0}{0}$ is not defined.)*
3. **Task** Let $\mathcal{D}$ be a dataset with $|\mathcal{D}| = n$, two different classes ($|\mathcal{Y}| = 2$) and $|\{(x_i, y_i) \in \mathcal{D} | y_i = \textit{positive}\}| = m$.
   - What's the expected F_1-score ($\mathbb{E}[F_1]$) of a classifier that uniformly picks a class at random?
   - What's the expected F_1-score of the former with $n - m = m$?

Confusion Matrices

Considering confusion matrices of the form

$$\begin{pmatrix} \text{True Positives (TP)} & \text{False Negatives (FN)} \\ \text{False Positives (FP)} & \text{True Negatives (TN)} \end{pmatrix}.$$

Based on the following confusion matrices, which classifier would you choose? Justify your answer since there might not be a single valid decision in some scenarios. Would you choose the classifier with confusion matrix C_1 or the classifier with C_2 ...



(a) Cats and Dogs **(b)** Trees **(c)** Fire Detection **(d)** Mushroom Soup

Figure 1 Imagine the following scenarios for a machine-learned classifier. What is a desireable result?

1. **Task** ...for distinguishing cats and dogs?

$$C_1 = \begin{pmatrix} 250 & 0 \\ 0 & 250 \end{pmatrix} \qquad C_2 = \begin{pmatrix} 125 & 125 \\ 125 & 125 \end{pmatrix}$$

2. **Task** ...for classifying a tree from its leafs?

$$C_1 = \begin{pmatrix} 0 & 250 \\ 250 & 0 \end{pmatrix} \qquad C_2 = \begin{pmatrix} 125 & 125 \\ 125 & 125 \end{pmatrix}$$

3. **Task** ...for fire detection task? (positive class means fire present)

$$C_1 = \begin{pmatrix} 200 & 25 \\ 75 & 200 \end{pmatrix} \qquad C_2 = \begin{pmatrix} 200 & 75 \\ 25 & 200 \end{pmatrix}$$

4. **Task** . . . for deciding which mushrooms in the forest to pick for your soup? (positive class means edible)

$$C_1 = \begin{pmatrix} 200 & 25 \\ 75 & 200 \end{pmatrix} \qquad C_2 = \begin{pmatrix} 200 & 75 \\ 25 & 200 \end{pmatrix}$$

Task 4: Extended Confusion Matrices

Given a classifier predicts the following $\hat{y}$ in a multi-class classification problem.

$\hat{y}$ (Predicted)	y (Actual Label)
Apple	Apple
Apple	Apple
Orange	Grapes
Grapes	Grapes
Orange	Orange
Apple	Grapes
Orange	Orange
Grapes	Orange
Grapes	Grapes
Grapes	Grapes
Orange	Apple
Grapes	Grapes
Apple	Apple
Orange	Orange
Apple	Apple
Orange	Apple
Grapes	Grapes
Grapes	Grapes
Orange	Orange
Orange	Orange
Grapes	Apple
Grapes	Grapes
Apple	Orange
Orange	Orange
Apple	Apple

1. **Task** Fill in the confusion matrix.

		Predicted		
		apple	**grapes**	**oranges**
Actual	**apple**			
	grapes			
	oranges			

2. **Task** Fill in the precision, recall, and F_1-score per class.

		Metrics		
		Precision	Recall	F_1
Actual	**apple**			
	grapes			
	oranges			

3. **Task** What's the overall F_1-score for the classifier with *micro-averaging*?
4. **Task** What's the overall F_1-score for the classifier with *macro-averaging*?