Problem Set 1

Oke

CEE 616: Probabilistic Machine Learning *09.11.2025*

Due Thursday, September 25, 2025 at 11:59PM on Canvas.

The problems are worth a total of **85 points**.

Programming requirements

It is recommended that you have a working installation of JupyterLab. Google Colab is also another viable notebook option. However, notebooks are optional, and you can always choose to write your code in an editor (e.g. Sublime Text) and process your outputs accordingly.

In Jupyter, you may use a Python, R or even MATLAB kernel.[1] Text can be formatted using Markdown (brief guide here: `https://learn.getgrav.org/content/markdown`).

Any Python packages you require must be installed locally to access them via the Jupyter kernel. (Use `conda` or `pip`.)

Programming help

I will provide some programming templates in Python/R in the coming days on Moodle, in order to ease some of the possible issues that may arise as you begin scripting. We will also go over coding examples weekly.

Submission instructions

There are two options for submission:

1. JupyterLab Notebook. Please name your notebook as follows:
   `<lastname>-<firstname>-PS1.ipynb`

2. R/Python/MATLAB script *and* PDF document with supporting responses. Your PDF should have complete responses to all the questions (including all the required plots). Your script should be clearly commented, producing all the results and plots you show in your PDF document. The filenames should be in a similar format as described above.

[1] In order to use the R kernel in Jupyter, you may have to install it, as well. Please follow the brief instructions here: `https://irkernel.github.io/installation/`. To install the MATLAB kernel, see `https://am111.readthedocs.io/en/latest/jmatlab_install.html`.

Problem 1 *True/False questions (10 points)*

Respond "T" (*True*) or "F" (*False*) to the following statements. Use the boxes provided. Each response is worth 1 point.

(i) ☐ Supervised learning only constitutes modeling frameworks that are suitable for regression.

(ii) ☐ Bayes' Theorem specifies a posterior probability $P(X|A)$ as inversely proportional to the product of its likelihood $P(A|X)$ and prior $P(X)$.

(iii) ☐ In the gradient descent optimization method, the second derivative of the loss function is not required to compute the update step.

(iv) ☐ Uncorrelated random variables are always independent of each other.

(v) ☐ In assessing a classifier, the receiver operating characteristics (ROC) curve is generated by plotting sensitivity versus specificity for different threshold values.

(vi) ☐ The determinant of a matrix can be given as the product of its eigenvalues.

(vii) ☐ The inverse of a matrix only exists if the matrix is singular.

(viii) ☐ An estimator is biased if the difference between its expectation and true value is zero.

(ix) ☐ Leave-one-out-cross-validation (LOOCV) can be considered a special case of k-fold CV in which $k = n$, where n is the number of observations in the dataset.

(x) ☐ In bootstrapping, each observation x_i in the original dataset is equally likely to be randomly selected for each resampled point in the sample.

Problem 2 *Bayes' theorem (9 pts)*

Given that $P(A) = 0.6$, $P(B) = 0.3$ and $P(C) = 0.1$ represent the production of machines in a factory. The conditional probabilities of defective items are $P(D|A) = 0.02$, $P(D|B) = 0.03$ and $P(D|C) = 0.04$.

(a) Find the total probability $P(D)$. *[3]*

(b) Find the probability that an item was produced by machine A, given that it is defective. *[3]*

(c) Draw a Venn diagram depicting the interaction among the events A, B, C and D in sample space S. *[3]*

Problem 3 *Estimating a linear model (16 pts)*

Given a vector of predictor variable samples $\boldsymbol{x}^T = \begin{bmatrix} 10 & 5 & 7 & 19 & 11 & 8 \end{bmatrix}$ and corresponding response vector $\boldsymbol{y}^T = \begin{bmatrix} 15 & 9 & 3 & 25 & 7 & 13 \end{bmatrix}$, the classical linear regression model can be written as

$$\hat{\boldsymbol{y}} = \boldsymbol{X}\hat{\boldsymbol{w}}$$

(a) Write the design matrix $\boldsymbol{X}$ in full assuming the model has an intercept (hint: the matrix should have two columns, the first being a column of 1's). *[2]*

(b) Using OLS (ordinary least squares) assumptions (see equation 4.59 on page 113 in **PMLI**), find $\hat{\boldsymbol{w}}$. (Show your work as much as possible, but the matrix multiplication can be done in Python/R/MATLAB. Include the code used if you are not submitting a Jupyter notebook.) *[4]*

(c) Find the vector of predicted values $\hat{\boldsymbol{y}}$. *[2]*

(d) Find the vector of residuals $\boldsymbol{e}$. *[2]*

(e) Compute the mean squared error (MSE) of this model. *[2]*

(f) Compute the root mean squared error (RMSE) of the model. *[1]*

(g) Create a scatterplot of the data and show the least squares line in the plot. *[3]*

Problem 4 *Logistic function (6 pts)*

The logistic sigmoid function is given by:

$$\boldsymbol{\sigma}(z) := \frac{1}{1 + e^{-z}} \tag{1}$$

(a) Produce a plot (in Python/R/Jupyter) of the function in the domain $z \in [-5, 5]$. *[2]*

(b) Show that its derivative is given by: *[4]*

$$\boldsymbol{\sigma}'(z) = \frac{d\boldsymbol{\sigma}(z)}{dz} = [1 - \boldsymbol{\sigma}(z)]\boldsymbol{\sigma}(z) \tag{2}$$

Problem 5 *Classifier performance (14 points)*

An estimated classification model produces the following confusion matrix on a test set:

		Predicted		
		Class 0	Class 1	Total
Observed	Class 0	9650	17	?
	Class 1	265	68	?
	Total	?	?	

[1] **(a)** What is the number of false positives (FP)?

[1] **(b)** What is the number of false negatives (FN)?

[1] **(c)** What is the number of positive observations (P)?

[1] **(d)** What is the number of predicted positive observations (P^*)?

[2] **(e)** Compute the test precision of the classifier.

[2] **(f)** Compute the test recall of the classifier.

[2] **(g)** Compute the test F_1-score of the classifier.

[2] **(h)** Compute the test accuracy of the classifier.

Problem 6 *Entropy (12 pts)*

PMLI Exercise 6.3 (page 218).

Problem 7 *Eigenvectors (6 pts)*

PMLI Exercise 7.2 (page 266). Note: you can check your answer with Python/R/Matlab (include the code you used).

Problem 8 *Linear system of equations (5 pts)*

Reformulate the two equations:

$$2x_1 + 6x_2 = 8$$
$$5x_1 + x_2 = 0$$

as a system of linear equations, and solve it for $[x_1\ x_2]^T$ using linear algebra.

Problem 9 *Subgradients (7 pts)*

PMLI Exercise 8.1 (page 314). [Context: the hinge loss is typically used as a loss function in classifier training, for example in support vector machines.] **Important:** Answer the question along the following steps:

(a) Sketch/plot the hinge loss function $f(x)$. *[2]*

(b) Write the piecewise subgradient/subderivative $\partial f(x)$ (see equation 8.14 on p. 276 for an example). *[2]*

(c) Evaluate the subgradient at the specified points in the exercise: $\partial f(0)$, $\partial f(1)$, $\partial f(2)$. *[3]*