# CS 6140: Machine Learning — Fall 2021— Paul Hand

Midterm 1
Due: Thursday October 28, 2021 at Noon Eastern time via Gradescope.

Name:

There are a total of 110 points available. The exam will be graded out of 100 points.

You must complete this exam by yourself. You may consult any and all resources other than people. Make sure to justify your answers. You may use a computer to perform calculations, such as linear algebra. You may either write your responses in LaTeX or you may write them by hand and take a photograph of them. You are encouraged to use Overleaf. Create a new project and replace the tex code with the tex file of this document, which you can find on the course website. When you upload your solutions to Gradescope, make sure to take each problem with the correct page or image.

**Question 1.**

(20 points) Consider the following training data.

| x | y |
|---|---|
| 0 | 0 |
| 1 | 1 |
| -1 | 2 |
| 2 | 3 |

Suppose the data comes from a model $y = c_0 + c_1 x + c_2 x^2$+noise for unknown constants $c_0, c_1, c_2$. Use the Normal Equations for least squares linear regression to find an estimate of $c_0, c_1, c_2$.

**Response:**

**Question 2.**

(25 points) Suppose you are given data $\{x_i, y_i\}_{i=1\ldots n}$ and weights $\{c_i\}_{i=1\ldots n}$, where $x_i \in \mathbb{R}^d$, $y_i \in \mathbb{R}$, and $c_i > 0$ for all $i$. Consider the following optimization problem:

$$\hat{\theta} = \arg\min_{\theta \in \mathbb{R}^d} \sum_{i=1}^{n} c_i (y_i - x_i^t \theta)^2.$$

Show that the solution to this problem is given by $(X^t D X)^{-1} X^t D y$, where $X$ is the $n \times d$ matrix whose $i$th row is $x_i^t$, $D$ is the $n \times n$ diagonal matrix where $D_{ii} = c_i$, and $y$ is the column vector whose $i$th entry is $y_i$. You may assume that the matrix $X^t D X$ is invertible.

**Response:**

**Question 3.**

Consider a binary classification problem over classes 0 and 1, where the data distribution is the following:

$$y_i = \begin{cases} 1 & \text{with probability } 1/2 \\ 0 & \text{with probability } 1/2, \end{cases} \qquad x_i \sim \begin{cases} \text{Uniform}([-1,2]) & \text{if } y_i = 1 \\ \text{Uniform}([-2,1]) & \text{if } y_1 = 0. \end{cases}$$

Note that the feature vector, $x$, is one dimensional, and the class label $y$ is either 0 or 1. Consider the following predictor: *If $x \geq t$, predict class 1; otherwise, predict class 0.*

(a) (10 points) Analytically compute and plot the recall of this predictor as a function of t. Express your answer in the form of a piecewise function of the following form

$$\text{recall}(t) = \begin{cases} \text{value} & -2 \leq t < -1, \\ \text{value} & -1 \leq t < 1, \\ \text{value} & 1 \leq t \leq 2. \end{cases}$$

**Response:**

(b) (10 points) Analytically compute and plot the precision of this predictor as a function of t. Express your answer in the same form as above. Hint: use Bayes theorem.

**Response:**

(c) (5 points) Plot the precision-recall curve.

**Response:**

**Question 4.**

(20 points) Consider logistic regression (with no bias term) over one-dimensional features. That is, consider a model where $P(\text{class } 1) = \sigma(\theta x)$, where $\sigma$ is the logistic function, $\theta \in \mathbb{R}$, and $x \in \mathbb{R}$. Consider the following data:

| x | y |
|---|---|
| -1 | 1 |
| 1 | 0 |

Write down the optimization problem given by logistic regression and express it in the form

$$\arg\min_{\theta \in \mathbb{R}} f(\theta).$$

Find and plot the function $f$.

**Response:**

**Question 5.**

You are solving a regression problem using linear regression. When you evaluate your model, you get a low training error but a high test error.

(a) (10 points)
Which of the following would be a better approach to help your model generalize better:

*Add additional features to the data set* OR *remove features from the dataset*?

Explain.

**Response:**

(b) (10 points)
Which of the following would be a better approach to help your model generalize better:

*Add more training examples* OR *remove training examples*?

Explain.

**Response:**