Machine Learning Course
Fall 2025

Westlake University
Department of Artificial Intelligence, SOE
Tao Lin
https://github.com/LINs-lab/course_machine_learning

Final Exam

Student Name: ______________________________ Student ID: ______________________________

Rules:

- This exam counts 40 percent towards your final grade.
- You have 180 minutes (from 14:00 till 17:00) to complete the exam.
- This is a closed book exam.
- Place on your desk: your student ID, one double-sided handwritten A4 page of summary if you have one, highly sugary drinks and ... if you need them.
- Place all your other personal belongings at the entrance or under your desk
- Write all your answers on the provided space. If you need extra paper let us know.

Good Luck!

1 Multiple Choice Questions [25 pts]

1. (XXX pts) Which of the following statements is true about the logistic regression model?

   (a) ☐ Logistic regression gives a max-margin classifier

   (b) ☐ By minimizing negative log-likelihood, we can obtain a closed-form solution for logistic regression

   (c) ☐ In logistic regression, we calculate the weights $\hat{\theta} = (X^T X)^{-1} X^T y$, and then fit responses as $\hat{y} = \sigma(X\hat{\theta})$

   (d) ☐ If we run Gradient Descent to solve a logistic regression task on linearly separable data, the weights will not converge

2 Multiple-Output Regression [25 pts]

Let $S = \{(\mathbf{y}_n, \mathbf{x}_n)\}_{n=1}^N$ be our training set for a regression problem with $\mathbf{x}_n \in^D$ as usual. But now $\mathbf{y}_n \in^K$, i.e., we have K outputs for each input. We want to fit a linear model for each of the K outputs, i.e., we now have K regressors $f_k(\cdot)$ of the form

$$f_k(\mathbf{x}) = \mathbf{x}^\top \mathbf{w}_k,$$

where each $\mathbf{w}_k^\top = (w_{k1}, \cdots, w_{kD})$ is the weight vector corresponding to the k-th regressor. Let $\mathbf{W}$ be the $D \times K$ matrix whose columns are the vectors $\mathbf{w}_k$.

Our goal is to minimize the following cost function $\mathcal{L}$:

$$\mathcal{L}(\mathbf{W}) = \sum_{k=1}^{K} \sum_{n=1}^{N} \frac{1}{2\sigma_k^2} (y_{nk} - \mathbf{x}_n^\top \mathbf{w}_k)^2 + \frac{1}{2} \sum_{k=1}^{K} \|\mathbf{w}_k\|_2^2,$$

where the σ_k are known real-valued scalars. Let $\boldsymbol{\sigma} = (\sigma_1, \cdots, \sigma_K)$.

For the solution, let $\mathbf{X}$ be the $N \times D$ matrix whose rows are the feature vectors $\mathbf{x}_n$.

1. (XXX pts) Write down the normal equations for $\mathbf{W}^\star$, the minimizer of the cost function. I.e., what is the first-order condition that $\mathbf{W}^\star$ has to fulfill in order to minimize $\mathcal{L}(\mathbf{W})$.

2. (XXX pts) Is the minimum $\mathbf{W}^\star$ unique? Assuming it is, write down an expression for this unique solution.

3. (XXX pts) Write down a probabilistic model, so that the MAP solution for this model coincides with minimizing the above cost function. Note that this will involve specifying the the likelihoods as well as a suitable prior (which will give you the regression term).