CS 189: Introduction to Machine Learning - Discussion 4

Questions to be discussed in section:

1. Intercepts in Linear Regression

   In the traditional linear regression scenario, where we model y with a line, or,

   $$\hat{y} = \vec{w}^T\vec{x}$$

   we aim to estimate $\vec{w}$. However, this model forces the lines to cross the origin (plug in $\vec{x} = \vec{0}$), severely limiting the power of the model. A typical solution to this problem is that the weight vector is extended by 1 and each input vector $\vec{x}$ has a 1 added at the beginning. This effectively adds an intercept term the model and allows for any line to be created.

   But that's boring! Let's come up with a solution to the intercept issue.

   Let's say we're stuck with the technique that fits only lines that go through the origin. We could shift the data to center it at zero, fit a line, and then shift our zero-centered line to where it's supposed to be. To center our data around the origin, we can subtract the mean of the x's and the mean of the y's from the data.

   a) Given that $\bar{x}$ and $\bar{y}$ are the means of our data, find the new model for a line predicting y from $\vec{x}$.

   Another approach to the intercept term would be to just model an intercept in our equation and estimate it from the data. The model would now look like this:

   $$\hat{y} = \vec{w}^T\vec{x} + w_0$$

   b) Find the MLE estimate of w_0. You should get the same answer as before. Assume that y is modeled with a line plus an intercept and Gaussian noise, i.e.,

   $$y \sim \mathcal{N}(\vec{w}^T\vec{x} + w_0, \sigma^2)$$

2. Linearly Separable Data with Logistic Regression

   Show (or explain) that for a linearly separable data set, the maximum likelihood solution for the logistic regression model is obtained by finding a vector β whose decision boundary $\beta^T x = 0$ separates the classes, and taking the magnitude of β to be infinity. **Note**: Remember that as mentioned in lecture, doing maximum-likelihood on logistic regression is same as minimizing cross-entropy loss (see lecture-6, slides-21,22). In lecture, we explored the cross-entropy loss-minimization perspective to logistic regression. This question will make you explore the likelihood perspective.

3. LMS algorithm

   Derive the LMS algorithm for a given dataset contaning m example pairs $(\mathbf{x}_i, y_i)$ where $\mathbf{x}_i \in \mathbb{R}^n$ and $y \in \mathbb{R}$. The model is assumed to be linear, i.e., $f(\mathbf{x}) = \mathbf{w}^T\mathbf{x}$. The aim is to minimize the squared loss over all examples and obtain the Stochastic Gradient Descent update equation.

Extra questions for practice:

4. Linear Algebra

   a) Let A be a square matrix. Show that we can write A as the sum of a symmetric matrix B and an antisymmetric matrix C:

   $$A = B + C$$

   (where $B = B^\top$ and $C = -C^\top$).

   b) Show that if C is antisymmetric, then $x^\top Cx = 0$ for all nonzero x.

   c) Show that the inverse of a symmetric matrix is symmetric.

   d) (Extra for Experts) Explain why covariance matrices of multivariate Gaussians can be assumed to be symmetric without loss of generality.