Lisa Yan
CS109

Section #8 Warmups

Based on the work of many CS109 staffs

1 Lecture 22, 5-27-20: Maximum A Posteriori

1. Intuitively, what is MAP? What problem is it trying to solve? How does it differ from MLE?
2. Given a 6-sided die (possibly unfair), you roll the die N times and observe the counts for each of the 6 outcomes as $n_1, ..., n_6$. What is the maximum a posteriori estimate of this distribution, using Laplace smoothing? Recall that the die rolls themselves follow a multinomial distribution.

2 Lecture 23, 5-29-20: Naive Bayes

Recall the classification setting: we have data vectors of the form $X = (X_1, \ldots, X_d)$ and we want to predict a label $Y \in \{0, 1\}$.

1. Recall in Naive Bayes, given a data point x, we compute $P(Y = 1|X = x)$ and predict $Y = 1$ provided this quantity is ≥ 0.5, and otherwise we predict $Y = 0$. Decompose $P(Y = 1|X = x)$ into smaller terms, and state where the Naive Bayes assumption is used.
2. Suppose we are given example vectors with labels provided. Give a formula to estimate (using maximum likelihood) each quantity $P(X_i = x_i|Y = y)$ above, for $i \in \{1, \ldots, d\}$ and $y \in \{0, 1\}$. You can assume there is a function count which takes in any number of boolean conditions and returns a count over the data of the number of examples in which they are true. For example, count$(X_3 = 2, X_5 = 7)$ returns the number of examples where $X_3 = 2$ and $X_5 = 7$.

3 Lecture 24, 6-1-20: Gradient Ascent and Linear Regression

Let $f : \mathbb{R}^n \rightarrow \mathbb{R}$ be a function which maps vectors $x \in \mathbb{R}^n$ to scalars $f(x) \in \mathbb{R}$.

1. What is the gradient ascent update step, with learning rate η?
2. Intuitively, what problem is gradient ascent trying to solve numerically?
3. What are some tradeoffs between a high and low learning rate (η)?

4 Lecture 25, 6-3-20: Logistic Regression

1. In general, how would we estimate the parameters for a model? For example, how would we estimate $\theta_0, \theta_1, \ldots, \theta_n$ for logistic regression?
2. Given parameters θ and a new sample x, how do we predict $\hat{y}$, i.e. the label for x? For now, assume that we are using binary labels, though you will soon see that we can extend logistic regression to a multiclass setting.