COEN 140 Machine Learning and Data Mining

Homework 1 (100 points)

Due: 12:10pm, Thursday, January 18, 2018

Submit a hardcopy of your solutions. You are welcome to type your solutions in LaTex. Writing your solutions by hand is also fine as long as they are neat.

Question 1 (50 points)

You're a 49ers fan, and the team is six weeks into its season. The number touchdowns scored in each game so far are given below: [2, 4, 3, 0, 3, 6]. Let's call these scores $x_1, x_2, \ldots, x_6$. Based on your data, you'd like to build a model to understand how many touchdowns the 49ers are likely to score in their next game. You decide to model the number of touchdowns scored per game using a Poisson distribution. The Poisson distribution with parameter λ assigns every non-negative integer x = 0, 1, 2, . . . a probability given by

$$Poisson(x|\lambda) = e^{-\lambda}\frac{\lambda^x}{x!}$$

So, for example, if λ = 1.5, then the probability that the 49ers score 2 touchdowns in their next game is $e^{-1.5} \times \frac{1.5^2}{2!} \approx 0.25$. To check your understanding of the Poisson, make sure you have a sense of whether raising λ will mean more touchdowns in general, or fewer.

1. (40 points) Derive an expression for the maximum-likelihood estimate of the parameter λ governing the Poisson distribution, in terms of your touchdown counts $x_1, x_2, \ldots, x_6$.

2. (10 points) Given the touchdown counts, what is your numerical estimate of λ?

Question 2 (50 points)

It is the first day of spring, and a gloomy 100-day winter has finally ended. Over those 100 days, $n_{clear} = 41$ of them were clear, $n_{cloudy} = 23$ of them were cloudy (but not rainy), and the remaining $n_{rainy} = 36$ were rainy. Denote the probability that a given winter day next year will be clear, cloudy, and rainy by p_{clear}, p_{cloudy}, and p_{rainy}, respectively, and assume that next winter's weather will follow the same pattern as this year's. Please derive expressions for the maximum-likelihood estimates of p_{clear}, p_{cloudy}, and p_{rainy}, in terms of this winter's weather data.