# ICS 635 HW2

February 7, 2019

## Bayesian Reasoning

### Problem 1: True Detective

Two people have left traces of their own blood at the scene of a crime. A suspect, Oliver, is tested and found to have type O blood. The blood groups of the two traces are found to be of type O (a common type in the local population, having frequency 60%) and of type AB (a rare type, with frequency $1\%$). Do these data (the blood types found at the scene) give evidence in favor of the proposition that Oliver was one of the two people whose blood was found at the scene? [From McKay, 2003]

### Problem 2: A Bent Coin

Suppose we have a bent coin. We treat a single coin flip as a random variable $X \in 0, 1$ where 0,1 correspond to tails, heads respectively. We model $X$ as a sample from a Bernoulli distribution parameterized by $\theta \in [0, 1]$, written as $X \sim$ Bernoulli($\theta$). Thus a particular value $x$ of random variable $X$ has probability

$$p(x) = \theta^x (1 - \theta)^{(1-x)}$$

Suppose we believe (subjectively) that any value of $\theta$ in the interval $[0, 1]$ is equally likely, and we can encode that information using a uniform prior. However, we will formulate the uniform prior as a Beta distribution $\theta \sim Beta(\alpha = 1, \beta = 1)$ for mathematical convenience (the Beta is the conjugate prior of the Bernoulli distribution, and is equivalent to the uniform prior when $\alpha = 1, \beta = 1$). The probability density function (pdf) of the Beta distribution is given by

$$p(\theta) = \frac{1}{Z(\alpha, \beta)} \theta^{\alpha-1} (1 - \theta)^{\beta-1}$$

The *partition* function $Z$ is a normalization factor that makes the pdf integrate to one, I.E. $\int_0^1 d\theta p(\theta) = 1$. For the Beta distribution, this $Z$ function makes use of the Gamma function $\Gamma$ which generalizes the factorial to real numbers.

$$Z(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha + \beta)}$$

You flip the coin once and it comes up heads.

1. What is the maximum likelihood estimate (MLE) of $\theta$?

2. What is the maximum a posterior (MAP) estimate of $\theta$?

3. What is the mean posterior (MP) estimate of $\theta$?

4. What is the posterior of $\theta$, $p(\theta|x)$?

5. What is the probability that the next flip will be heads, $p(x_2 = 1|x = 1, \theta)$?

6. You flip the original coin a total of 5 times and it comes up heads every time. Now what are the MLE, MAP, and MP estimates for $\theta$?

7. Setting the beta prior: Suppose we obtain a new bent coin, and an examination leads us to believe that $E[\theta] = m$ and $var[\theta] = v$. What values of $\alpha$ and $\beta$ will give our prior these properties? [From Murphy 2012]

### Problem 3: Naive Bayes by Hand

Consider a Naive Bayes model (multivariate Bernoulli version) for spam classification with the vocabulary $V = \{secret, offer, low, price, valued, customer, today, dollar, million, sports, is, for, play, healthy, pizza\}$. We have the following example spam messages, "*million dollar offer*", "*secret offer today*", "*secret is secret*" and normal messages, "*low price for valued customer*", "*play secret sports today*", "*sports is healthy*", "*low price pizza*". Give the MLEs for the following parameters: $\theta_{spam}$, $\theta_{secret|spam}$, $\theta_{secret|non-spam}$, $\theta_{sports|non-spam}$, $\theta_{dollar|spam}$. [From Daphne Koller]

## Supervised Learning

### Problem 4: Train vs. Test Error

1. The error on the test will always *decrease* as we get more training data, since the model will be better estimated. However, for complex models, the error on the training set can *increase* as we get more training data, until we reach some plateau. Explain why.

2. Suppose we have a randomly labeled dataset (i.e., the features x tell us nothing about the class labels y) with $N_1$ examples of class 1, and $N_2$ examples of class 2, where $N_1 = N_2$. What is the best misclassification rate any method can achieve? What is the estimated misclassification rate of the same method using Leave-One-Out-Cross-Validation? [From Witten05, p152.]

## Problem 5: Quasars

Classify quasars from galaxies and stars, using real data from the Sloan Digital Sky Survey. The following Colab notebook provides example code for a 1-NearestNeighbor classifier. Modify the code to improve the test set classification accuracy using *each* of the following classifiers: KNN, Gaussian Naive Bayes, Linear Discriminant Analysis, Quadratic Discriminant Analysis, and Decision Trees. Try to get good results for each by tuning the hyperparameters. You don't need to turn in code, but describe for each model (1) the data preprocessing and hyperparameters used and how you chose them, and (2) the expected generalization error, providing enough details to reproduce your results.

**Note: Over-estimating generalization performance will be penalized, since this is a common mistake in machine learning.**

`https://github.com/peterjsadowski/sklearn_examples/blob/master/sdss/quasars.ipynb`