Assignment 3

Problem 1

Suppose we flip a coin N times and observe N_0 tails and N_1 heads. We believe the outcome of a flip is determined by a Bernoulli with parameter θ.

1. Let $\hat{\theta}_1$ be the MAP under a uniform prior (i.e. $P(\theta) = 1$). Derive $\hat{\theta}_1$.
2. Consider the following prior for θ that believes the coin is either fair or slightly biased towards tails.

$$p(\theta) = \begin{cases} 0.5 & \text{if } \theta = 0.5 \\ 0.5 & \text{if } \theta = 0.4 \\ 0 & \text{otherwise} \end{cases}$$

   Derive the MAP estimate of θ under this prior, $\hat{\theta}_2$.

3. Suppose the true parameter is $\theta = 0.41$, and suppose $N = 10$. What is the probability that $\hat{\theta}_2$ is closer to the true value than $\hat{\theta}_1$? That is, what is

$$P(|\hat{\theta}_2 - \theta| < |\hat{\theta}_1 - \theta|)?$$

   You may find the following binomial distribution table for $\theta = 0.41$ helpful.

Binomial distribution (n=10, p=0.41)			
	f(x)	F(x)	1 - F(x)
x	Pr[$X = x$]	Pr[$X \leq x$]	Pr[$X > x$]
0	0.0051	0.0051	0.9949
1	0.0355	0.0406	0.9594
2	0.1111	0.1517	0.8483
3	0.2058	0.3575	0.6425
4	0.2503	0.6078	0.3922
5	0.2087	0.8166	0.1834
6	0.1209	0.9374	0.0626
7	0.0480	0.9854	0.0146
8	0.0125	0.9979	0.0021
9	0.0019	0.9999	0.0001
10	0.0001	1.0000	-0.0000

4. Suppose instead $N = 10000$. Which estimator do you think will usually be close to the true value? Why?

Problem 2

Suppose we have 3 different trained neural network models that we are considering using for predicting whether a photo (a matrix of 10,000 by 10,000 black/white intensity pixel values) contains a dog or a cat. Each neural network takes a photo as input and outputs a probability for dog and cat respectively. We also have 1000 examples of labeled dogs and cats from the Asirra data set. In addition, we have a new picture (of either a dog or a cat) and we want to know if it is a cat.

In the questions below, we are asking you to translate concepts in English into formal notation. You should use notation such that someone knowledgeable about statistical machine learning can tell exactly what your notation specifies with no uncertainty. You do not need to derive solutions or show how to compute an expression.

- Define the variables and expressions you will need in for the questions below.
- Which model do we think is best?
- If we assume that these three models are the only possibilities for how the labels were generated, what do we think is probability that model 1 is the true model?
- What does model 1 predict is the probability that the new picture depicts a cat?
- Without having made a hard choice of model, what do we think is the probability that the new picture depicts a cat?

Problem 3

Suppose we have two sensors with known (and different) variances v_1 and v_2, but unknown (and the same) mean μ. Suppose we observe n_1 observations $y_i^{(1)} \sim \mathcal{N}(\mu, v_1)$ from the first sensor and n_2 observations $y_i^{(2)} \sim \mathcal{N}(\mu, v_2)$ from the second sensor. (For example, suppose μ is the true temperature outside, and sensor 1 is a precise (low variance) digital thermosensing device, and sensor 2 is an imprecise (high variance) mercury thermometer.)

Let D represent all the data from both sensors. What is the posterior $p(\mu|D)$, assuming a non-informative prior for μ (which we can simulate using a Gaussian with a variance of ∞)? Give an explicit expression for the posterior mean and variance.

Problem 4 (optional)

You are investigating the relationship between students' study hours and their exam scores. You suspect that there is an unobserved latent variable, the students' innate ability, which influences both their study hours and exam scores. You model the latent ability as a Gaussian variable with a mean of 0 and a standard deviation of 1. The study hours and exam scores are also modeled as Gaussian variables, given the latent ability.

The study hours, given the latent ability, follow a Gaussian distribution with a mean of 2 times the latent ability and a standard deviation of 1.5. The exam scores, given the latent ability, follow a Gaussian distribution with a mean of 3 times the latent ability and a standard deviation of 2.

Assignment 3

If a student's exam score is observed to be 75, and the student spent 30 hours studying, use Bayes' rule to estimate the posterior distribution of the latent ability.

- Write down the expressions for the likelihood functions, prior probabilities, and posterior probabilities in this context.
- Calculate the posterior distribution of the latent ability given the observed exam score and study hours.