CSE 515T (Spring 2015) Midterm

- There are two ways to hand in this midterm. **Late submissions will not be accepted!** I do not recommend cutting it too close.
  - Physically in my department mailbox in the CSE head office on the fifth floor of Bryan Hall. The due date for this option is **5:00 PM, Friday, 6 March.**
  - Electronically on Piazza as a private message to the instructors. The due date for this option is **midnight, Saturday, 7 March.**
- Please do not discuss the questions with other members of the class.
- Please post any questions as a *private message to the instructors* on Piazza.
- Any corrections will be posted by the instructors on Piazza. This document will also be kept up-to-date on the course webpage and in GitHub.

1. Consider two coins with unknown bias θ_1 and θ_2, respectively. We place independent, identical beta priors on these quantities:

$$p(\theta_1) = \mathcal{B}(\theta_1; 2, 2); \qquad p(\theta_2) = \mathcal{B}(\theta_2; 2, 2).$$

Imagine someone flips both coins and tells you that *exactly one* of the outcomes (but not which) was a "head." Thus the observation was either HT or TH, but you are not told which. The below expressions are conditioned on "H" to indicate this observation.

- Given an expression for the posterior of the first coin's bias given this observation, $p(\theta_1 \mid \mathrm{H})$. Simplify the result as much as you can. Plot the prior and the posterior for θ_1 over the interval $\theta_1 \in (0, 1)$.
- Give an expression for the joint posterior $p(\theta_1, \theta_2 \mid \mathrm{H})$. Plot the joint prior, the likelihood, and the joint posterior as three separate heat maps over the unit square $(\theta_1, \theta_2) \in (0, 1)^2$. Use a grid with at least 100 values along each of the two θ axes.
- Summarize what the observation taught us about the bias of the coins.

2. Consider the three-dimensional parameter vector $\boldsymbol{\theta} = [\theta_1, \theta_2, \theta_3]^\top$, with the following joint multivariate Gaussian prior:

$$p(\boldsymbol{\theta}) = \mathcal{N}(\boldsymbol{\theta}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \mathcal{N}\left(\begin{bmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \end{bmatrix}; \begin{bmatrix} 0 \\ 1 \\ 2 \end{bmatrix}, \begin{bmatrix} 1 & 2 & 0 \\ 2 & 9 & 0 \\ 0 & 0 & 16 \end{bmatrix} \right).$$

We are going to consider a decision problem with action space $\mathcal{A} = \{1, 2, 3\}$. The result of choosing an action $a \in \mathcal{A}$ will be to observe the exact value of θ_a, the ath element of $\boldsymbol{\theta}$.

Consider the following loss functions, ℓ_1 and ℓ_2:

$$\ell_1(\boldsymbol{\theta}, a) = \begin{cases} 1 & \theta_a > 0 \\ 0 & \theta_a \leq 0 \end{cases} \qquad \ell_2(\boldsymbol{\theta}, a) = \min(0, \theta_a).$$

For each:

- Write a generic expression for the expected loss of action a in terms of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$. Evaluate any integrals you encounter.
- Give a numerical value for the expected loss of each action, using the values of $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ provided above.
- State the Bayes action.

3. Consider a d-dimensional vector $\boldsymbol{\theta}$ with an arbitrary multivariate Gaussian distribution:

$$p(\boldsymbol{\theta}) = \mathcal{N}(\boldsymbol{\theta}; \boldsymbol{\mu}, \boldsymbol{\Sigma}).$$

- Give a general expression for the distribution of the following (scalar) value τ.

$$\tau = \theta_1 + 2\theta_2 + \cdots d\theta_d$$

- Consider again the specific distribution of the three-dimensional vector $\boldsymbol{\theta}$ from the last problem, as well as the action space $\mathcal{A}$ with the same observation mechanism: after choosing $a \in \mathcal{A}$, we will observe the corresponding value θ_a. Suppose we may select one action and then must predict τ under a squared loss function:

$$\ell(\tau, \hat{\tau}) = (\tau - \hat{\tau})^2.$$

Using the distribution from the last problem. what is the expected loss of each of the three available actions? Which is the Bayes action?

4. Consider the following data:

$$\mathbf{x} = [0.54, 1.84, -2.26, 0.86, 0.32]^\top;$$
$$\mathbf{y} = [-1.31, -0.43, 0.34, 3.58, 2.77]^\top.$$

Consider the Bayesian linear regression model with $\phi(x) = [1, x]^\top$. Use the prior $p(\mathbf{w}) = \mathcal{N}(\mathbf{w}; \mathbf{0}, \mathbf{I})$.

Plot the posterior probability that the slope of the regression line is positive as a function of the standard deviation of the observation noise σ (the noise variance is then σ^2). Use a grid of at least 100 points in the range $\sigma \in (0.01, 10)$.