Stats 545: Midterm exam

This is a 75-minute exam for 32 points. Write your name and PUID on each sheet, and also include the number of answer sheets. Attempt all questions.

1 Optimization [5 pts]

Let $f(x)$ be a continuous 1-dimensional function. Let x^* be a root of f, i.e. $f(x^*) = 0$.

1. Give a simple condition to verify that an interval $[a, b]$ contains x^*. [1pts]
2. Given such an initial bracketing $[a, b]$ that contains x^*, write a few lines of `R` to return a solution x_{soln} up to some accuracy `eps`, i.e. so that $|x^* - x_{soln}| <$ `eps`. Provide a few lines of pseudocode first if it helps (or for partial credit). [3pts]
3. Briefly explain the pros and cons of gradient descent vs EM for finding an MLE solution. [1pt]

2 Rejection sampling [6 pts]

Let $p(\theta) = \frac{1}{\pi(1+\theta^2)}$ (the Cauchy distribution), and $p(X|\theta) = \mathcal{N}(\theta, 1)$ (i.e. a Gaussian with mean θ and variance 1). You only have access to Gaussian and uniform random number generators.

1. Write down an expression for $p(X)$, the marginal probability of X, and briefly describe a Monte Carlo algorithm to approximate this. For partial credit, you can assume access to any random number generator. [3pts]
2. Write down an expression for $p(\theta|X)$, and briefly describe a rejection sampling algorithm to sample for this, giving both the proposal distribution and the acceptance probability. [3pts].

3 MCMC [7 pts]

1. Briefly explain the difference between Monte Carlo sampling and Markov chain Monte Carlo sampling. Explain when you might use one or the other. [2pts]
2. Briefly explain what burn-in is for MCMC. What is 'effective sample size'? [2pts]
3. What does irreducibility mean for a Markov chain? Suppose $p(x)$ is some distribution on the real line. Let the proposal distribution $q(x|x_{old})$ be a uniform distribution centered at x_{old} i.e. $q(x|x_{old}) = \text{Unif}(x_{old} - \frac{1}{2}, x_{old} + \frac{1}{2})$. Give examples of $p(x)$ for which the corresponding MH Markov chain is and is not irreducible, briefly explaining why. [3pts]

4 MCMC 2 [10 pts]

You want to sample from $p(x, y) \propto \exp\left(-x^2 - (y - x^2)^2\right)$ where both x and y are real-valued. The only random-number generators you have are a Gaussian and a uniform random number generator.

1. Describe a Metropolis-Hastings sampler to do this, giving the proposal distribution and the acceptance probability.[4pts]

Next, we will write down a Gibbs sampler to do this.

2. What is the conditional distribution $p(y|x)$? In particular, what family does it belong to? [1pts]
3. Write down the conditional distribution $p(x|y)$? That is not a standard distribution. If you were to sample from this by rejection sampling, write down the proposal distribution and acceptance probability. [2pts]
4. Describe the overall Gibbs sampling algorithm briefly, and how you would use it to calculate the mean of y. [2pts]
5. Suggest a few pros and cons of Gibbs sampling vs Metropolis-Hastings. [1pts]

5 k-nearest neighbors [4 pts]

1. What is overfitting? For a given dataset, is k-nearest neighbours with a large k or a small k more likely to cause overfitting. Explain why briefly. [2pts]
2. Explain briefly what crossvalidation is and how/why it avoids overfitting. [2pts]