CS534 — Homework Assignment 1 —
Due Oct 6th 11:59pm, 2018

Written assignment

Individual assignment. Submit electronically via TEACH (https://teach.engr.oregonstate.edu/).

1. (Probability) Consider two coins, one is fair and the other one has a 1/10 probability for head. Now you randomly pick one of the coins, and toss it twice. Answer the following questions.
   (a) What is the probability that you picked the fair coin? What is the probability of the first toss being head?
   (b) If both tosses are heads, what is the probability that you have chosen the fair coin (Hint: Bayes Rule)?

2. (Maximum likelihood estimation for uniform distribution.) Given a set of i.i.d. samples $x_1, x_2, ..., x_n \sim uniform(0, \theta)$.
   (a) Write down the likelihood function of θ.
   (b) Find the maximum likelihood estimator for θ.

3. (Weighted linear regression) In class when discussing linear regression, we assume that the Gaussian noise is independently identically distributed. Now we assume the noises $\epsilon_1, \epsilon_2, \cdots \epsilon_n$ are independent but each $\epsilon_m \sim N(0, \sigma_m^2)$, i.e., it has its own distinct variance.
   (a) Write down the log likelihood function of $\mathbf{w}$.
   (b) Show that maximizing the log likelihood is equivalent to minimizing a weighted least square loss function $J(\mathbf{W}) = \frac{1}{2}\sum_{m=1}^{n} a_m(\mathbf{w}^T\mathbf{x}_m - y_m)^2$, and express each a_m in terms of σ_m.
   (c) Derive a batch gradient descent algorithm for optimizing this objective.
   (d) Derive a closed form solution to this optimization problem.

4. (Decision theory). Consider a binary classification task with the following loss matrix:

predicted label $\hat{y}$	true label y 0	1
0	0	10
1	5	0

   We have build a probabilistic model that for each example x gives us an estimated $P(y = 1|x)$. It can be shown that, to minimize the expected loss for our decision, we should set a probability threshold θ and predict $\hat{y} = 1$ if $P(y = 1|x) > \theta$ and $\hat{y} = 0$ otherwise.
   (a) Please compute the θ for the above given loss matrix.
   (b) Show a loss matrix where the threshold is 0.1.

5. Consider the maximum likelihood estimation problem for multi-class logistic regression using the soft-max function defined below:

$$p(y = k|\mathbf{x}) = \frac{\exp(\mathbf{w}_k^T\mathbf{x})}{\sum_{j=1}^{K}\exp(\mathbf{w}_j^T\mathbf{x})}$$

   We can write out the likelihood function as:

$$L(\mathbf{w}) = \prod_{i=1}^{N}\prod_{k=1}^{K} p(y = k|\mathbf{x}_i)^{I(y_i=k)}$$

   where $I(y_i = k)$ is the indicator function, taking value 1 if y_i is k.

(a) What are i and k in this likelihood function?

(b) Compute the log-likelihood function.

(c) What is the gradient of the log-likelihood function w.r.t the weight vector $\mathbf{w}_c$ of class c? (Precursor to this question, which terms are relevant for $\mathbf{w}_c$ in the loglikelihood function?)