Stat Learn & Graph Models

HW 4

Due date: 11/28/2017

Total points: 100 + 15*
UG grading : $\min(p, 100)$
Grad grading : $\min(p, 110)/110 \times 100$
where p is the points obtained in the HW.

Problem 1 Classification with Poisson Class-conditionals

(15 pts, 5 pts/part) Consider a classification problem with

$$\begin{aligned} p(y=0) &= \pi_0 \\ p(y=1) &= \pi_1 \\ p(x|y=0) &= \frac{\lambda_0^x e^{-\lambda_0}}{x!} \\ p(x|y=1) &= \frac{\lambda_1^x e^{-\lambda_1}}{x!} \end{aligned}$$

where x is a nonnegative integer, y is 0/1-binary, and $\pi_0 + \pi_1 = 1$.

a) Assuming a dataset $\{(x_i, y_i)\}_{i=1}^N$, find the maximum likelihood estimates for λ_0 and λ_1.

b) Prove that the $p(y = 1|x)$ has the logistic form. That is $p(y = 1|x) = \frac{1}{1+e^{-\beta^T \tilde{x}}}$, where $\tilde{x} = (1, x)^T$ and $\beta = (\beta_0, \beta_1)^T$. Determine β_0 and β_1.

c) Determine the decision boundary.

Problem 2 The gradient for Logistic Regression

(10 pts) Consider the logistic regression model in which $p(y|x) = \mu(x)^y (1 - \mu(x))^{1-y}$, where $\mu(x) = \frac{1}{1+e^{-\beta^T x}}$. The goal is to find the ML estimate for β, based on the dataset $\{(x_i, y_i)\}_{i=1}^N$. Let $\mu_i = \mu(x_i)$. The log-likelihood function is

$$\ell(\beta) = \sum_{i=1}^N (y_i \ln \mu_i + (1 - y_i) \ln(1 - \mu_i))$$

Prove that $\nabla_\beta \ell(\beta) = \sum_{i=1}^N (y_i - \mu_i) x_i$.

Problem 3 EM vs. ML

(25 pts, 5 pts/part) (Capital letters denote random variables.) We saw in our in-class activity that if $X = S + N$, where $S \sim \mathcal{N}(0, \theta)$ and $N \sim \mathcal{N}(0, \sigma^2)$, with σ^2 known and $\theta > 0$ but unknown, the new EM estimate for θ when $X = x$ is

$$\theta'' = E[S^2 | X = x, \theta']$$

a) Argue that the pair (S, X) is jointly Gaussian. (Hint: Two random variables are jointly Gaussian if every linear combination of the two is Gaussian.)

b) Find the mean vector and the covariance matrix of this joint distribution.

c) Find the distribution for $S|X = x, \theta$, using the following fact. If Y and Z are zero-mean and jointly Gaussian, then $Y|Z = z$ has distribution $\mathcal{N}(a, b)$, where $a = \frac{Cov(Y,Z)}{Var(Z)}z$ and $b = Var(Y) - \frac{Cov(Y,Z)^2}{Var(Z)}$.

d) Let $f(\theta) = \left(\frac{\theta x}{\theta+\sigma^2}\right)^2 + \frac{\theta\sigma^2}{\theta+\sigma^2}$. Use the previous part to show that

$$E[S^2|X = x, \theta'] = f(\theta').$$

This means that $\theta'' = f(\theta')$.

e) Show that $\theta = x^2 - \sigma^2$ (the solution with direct ML approach) is a stationary point of $f(\theta)$, if $x^2 \geq \sigma^2$.

Problem 4 EM

(35 pts) Suppose that we have K coins numbered $1, \dots, K$. The probability of heads for coin k is τ_k. We perform the following N experiments. In experiment i, we pick a coin at random, where coin k is chosen with probability π_k. We don't actually know which coin we have picked (hidden data). We toss the coin and record $x_i = 1$ if head shows and $x_i = 0$ otherwise (observed data). Let $\theta = (\pi, \tau)$ determine the (unknown) parameter set. Find the EM recursion for θ.

Hint: Define the random vector Y_i that determines which coin is chosen in experiment i as a binary vector of length K. If coin k is chosen in experiment i, all coordinates of Y_i are zero except the kth coordinate, which is equal to 1.

Problem 5 Non-negativity of the Relative Entropy

a) (8 pts) The log-sum inequality states that for nonnegative numbers $a_1, \dots, a_n$ and $b_1, \dots, b_n$,

$$\sum_{i=1}^{n} a_i \log \frac{a_i}{b_i} \geq \left(\sum_{i=1}^{n} a_i\right) \log \frac{\sum_{i=1}^{n} a_i}{\sum_{i=1}^{n} b_i},$$

with equality if and only if $\frac{a_i}{b_i} = \text{const}$. Prove it using the Jensen inequality: For a convex function f and a random variable X, $f(EX) \leq Ef(X)$, and, moreover, if f is strictly convex, equality holds if and only if X is constant.

b) (7 pts) Using the log-sum inequality prove that $D(p||q) \geq 0$, for two distributions p and q, with equality if and only if $p \equiv q$.

Problem 6 MMSE Estimators

(15 pts*) (Random variables shown by capital letters.) Suppose that X, Y are random variables with the joint pdf

$$f(x, y) = \begin{cases} x + y, & 0 \leq x, y \leq 1, \\ 0, & else \end{cases}$$

Find the linear and unconstrained estimators for Y if $X = x$ is known.