Machine Learning: Homework Assignment 6

E4525 Spring 2018,

IEOR, Columbia University

Due:Apr 19th, 2018

1. **Exponential Distribution**

   $s \in \mathbb{R}^+$ (s continuous and $s > 0$) has exponential distribution

   $$p(s;\lambda) = \lambda e^{-\lambda s} \qquad \text{for} \quad \lambda > 0 \tag{1}$$

   which is an exponential family distribution.

   Give expressions for

   (a) Base measure $h(s)$.

   (b) Sufficient statistic $T(s)$.

   (c) canonical parameter η as a function of the natural parameter λ.

   (d) The cumulant function $A(\eta)$.

   (e) The average and variance of s as a function of λ

2. **Gamma Distribution, shape parameter α known**

   $s \in \mathbb{R}^+$ (s continuous and $s > 0$), s has Gamma distribution

   $$p(s;\beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} s^{\alpha-1} e^{-\beta s} \qquad \text{for} \quad \alpha > 0,\ \beta > 0 \tag{2}$$

   if we assume α is a known fixed number, $p(s;\beta)$ is has an exponential family distribution with a scalar canonical parameter η.

   Give expressions for

   (a) Base measure $h(s)$.

   (b) Sufficient statistic $T(s)$.

   (c) canonical parameter η as a function of the natural parameter β.

   (d) What does the constrain $\beta > 0$ in the definition of the Gamma distribution imply for the canonical parameter?

   (e) The cumulant function $A(\eta)$.

   (f) The average and variance of s as a function of α and β

(g) show that the exponential distribution of exercise 1 is an special case of the Gamma distribution.

3. **Gamma Distribution, shape parameter α unknown**

   $s \in \mathbb{R}^+$ (s continuous and $s > 0$) has distribution

$$p(s; \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} s^{\alpha-1} e^{-\beta s} \quad \text{for} \quad \alpha > 0,\ \beta > 0. \tag{3}$$

   If we assume α is unknown number, $p(s; \alpha, \beta)$ is an exponential family distribution with a two dimensional canonical parameter $\eta = (\eta_1, \eta_2)$.

   Give expressions for

   (a) Base measure $h(s)$.

   (b) Sufficient statistic $T(s) = (T_1(s), T_2(s))$.

   (c) canonical parameter $\eta = (\eta_1, \eta_2)$ as a function of the natural parameters (α, β).

   (d) The cumulant function $A(\eta)$.

4. **Generalized Linear Model for Gamma Distribution: canonical link**

   We have two random variables $X \in \mathbb{R}^D$ and $Y \in \mathbb{R}^+$. We assume that $Y|_X$ is a Gamma distribution with the same assumptions of problem 2: Y has a Gamma distribution with α fixed

$$p(y|x; \eta, \alpha) = h(y)e^{\eta y - A(\eta)} \tag{4}$$

   where $h(y)$ and $A(\eta)$ are given as the solutions of problem 2.

   We further assume the canonical link (note the minus sign)

$$\eta(x) = -(b + x^T w). \tag{5}$$

   where $w = (w_1, \ldots, w_D)$.

   Given N observations $\{x_i, y_i\}$ give expression for the following

   (a) The maximum likelihood loss function $E_{\text{Gam}}(b, w; \{x_i, y_i\})$

   (b) The expected value $\hat{y}(x; b, w)$ of y conditional on x as a function of α, b and w

   (c) The gradient of the loss E_{Gam} for changes of b and w_d

   (d) What does the constrain $\beta > 0$ in the definition of the Gamma probability distribution imply for the parameters b and w?

5. **Generalized Linear Model for Gamma Distribution: log link**

   To avoid the constrained optimization that you analyzed in problem 4d We have two random variables X and Y. We assume that $Y|_X$ is a Gamma distribution with the assumptions of problem 2 For $s \in \mathbb{R}^+$ (s continuous and $s > 0$) has Gamma distribution

$$p(y|x;\eta,\alpha) = h(s)e^{\eta s - A(s)} \tag{6}$$

where η, $h(s)$ and $A(\eta)$ are given as the solutions of problem 2.

We further a non-canonical link defined by $\psi(\eta) = -e^{-\eta}$

$$\eta(x) = \psi(b + x^T b) = -e^{-b-x^T w}. \tag{7}$$

Given N observations $\{x_i, y_i\}$ give expression for the following

(a) The maximum likelihood loss function $E^{\mathrm{log}}_{\mathrm{Gam}}(b, w; \{x_i, y_i\})$

(b) The expected value $\hat{y}(x; b, w)$ of y conditional on x as a function of α, b and w

(c) The gradient of the loss $E^{\mathrm{log}}_{\mathrm{Gam}}$ for changes of b and w_d