ANLY601 Assignment1

Mengtong Zhang

January 2020

Collaborate with Shaoyu Feng and Yunjia Zeng

1 Likelihood Estimatiion

1.1

$$L(\theta) = \theta^n (1-\theta)^{\sum_1^n (x_i) - n}$$
$$ln(L(\theta)) = nln(\theta) + (\sum_1^n x_i - n) ln(1-\theta)$$

Let the derivative=0,

$$\frac{n}{\theta} - \frac{\sum_1^n x_i - n}{1-\theta} = 0$$
$$\hat{\theta}_{MLE} = \frac{n}{\sum_1^n x_i}$$

1.2

$$L(a,b) = \frac{1}{(b-a)^n}$$
$$ln(L(a,b)) = -nln(b-a)$$

To make the likelihood function as large as possible, simply take

$$\hat{a}_{MLE} = min(x_i)$$
$$\hat{b}_{MLE} = max(x_i)$$

2 Loss Functions

2.1

Denote Loss Function as LL,

$$LL = \sum_{n=1}^{N} log(N(x_n|\mu, \sigma^2) = \sum_{n=1}^{N} log(\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\frac{(x_n-\mu)^2}{\sigma^2}})$$

$$LL = \sum_{n=1}^{N}(-log(\sqrt{2\pi\sigma^2}) + (-0.5)\frac{(x_n-\mu)^2}{\sigma^2})$$

$$LL = -\frac{N}{2}log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{n-1} N(x_n-\mu)^2$$

when we already know σ, LL is equivalent to the second term, which is equivalent to squared error loss.

2.2

$$LL = \sum_{n=1} Nlog(L(x_n|\mu,\theta)) = \sum_{n=1}^{N}\frac{1}{2\theta}e^{(}-\frac{|x-\mu|}{\theta}))$$

$$LL = -Nlog(2\theta) - \frac{1}{b}\sum_{n=1}^{N}|x-\mu|$$

when we already know θ, LL is equivalent to the second term, which is equivalent to mean absolute loss.

3 Decision Rules

3.1

$$L(\theta,\delta(X)) = E[(\theta-\delta(X))^2] = VAR(\delta(X)) + E(\delta(X)-\theta)^2$$

Because the decision rule is unbiased,

$$E(\delta(X)-\theta)^2 = 0$$

$$\delta(X) = \bar{X}$$

3.2

$$\min_{c\in R} E[|X-c|] = \min_{c\in R}\left(-0.5\int_{-\infty}^{c}(x-c)f(x)dx + 0.5\int_{c}^{\infty}(x-c)f(x)dx\right)$$

$$0 = 0.5\int_{-\infty}^{c}f(x)dx - 0.5\int_{c}^{\infty}f(x)dx = 0.5F(c) - 0.5(1-F(c)) = F(c) - 0.5$$

Take $c = median(x)$, we can make mean absolute error=0, which is optimal.

4 Convexity

4.1

$$\frac{dL}{d\beta} = \frac{dL}{dp}\frac{dp}{d\beta} = -(\frac{y}{p} - \frac{1-y}{1-p})(\frac{\beta e^{-\beta x}}{(e^{-\beta x}+1)^2})$$

$$\frac{d^2L}{d\beta^2} = \frac{x^2 e^{\beta x}}{(e^{\beta x}+1)^2} \geq 0$$

Thus the cross entropy loss is convex with respect to β.

4.2

$$\frac{dL}{d\beta} = \frac{dL}{dp} * \frac{dp}{d\beta} = -2(y-p)p(1-p)x$$

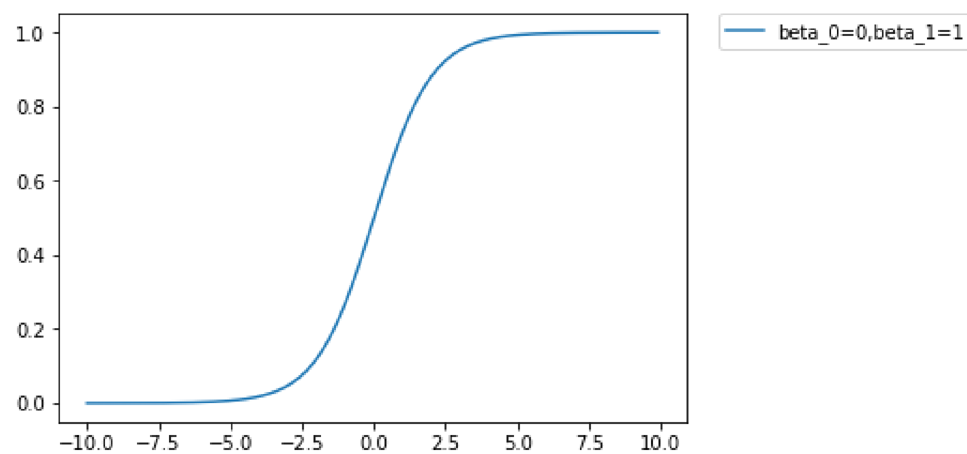
$$\frac{d^2L}{d\beta^2} = -2[y - 2yp - 2p + 3p^2]x^2 p(1-p)$$

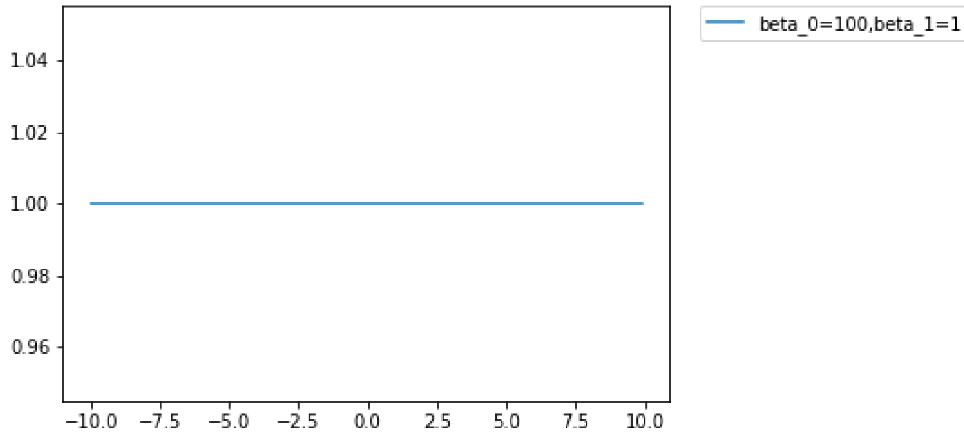
when $y = 0$, the second order derivative is positive only when p is in range $[0, 2/3]$, this disprove the convexity of Mean squared loss.

5 Decision Boundaries

5.1

$\theta = 0$, $f_\theta(0) = 1/(1 + exp(-\beta_0 - \beta_1 x)) = 1/2$. Thus, the decision boundary is $f_\theta(x) = 0.5$.
$\beta_0 = 100$, $f_\theta(0) = 1/(1+exp(-100)) = 1$, All points will be classified to class B.





5.2

$logit(f_\theta(x)) = log(f_\theta(x)/(1 - f_\theta(x))) = log(1/exp(-\beta x)) = \beta x$
Since *logit* is a monotonous function, suppose the decision boundary is $f_\theta(x) = c$, then it could also be written as $\beta x = logit^{-1}(c)$. Thus, $\theta \cdot x = \theta_0 + \theta_1 x$ is a linear separating hyperplane.

6 Sufficient Statistic

$$f(x_1, ..., x_n|\mu) = (2\pi)^{-n/2}\sigma^{-n}e^{(\frac{-1}{2\sigma^2}\sum_{i=1}^n (x_i-\mu)^2)}$$
$$= (2\pi)^{-n/2}\sigma^{-n}e^{(\frac{-1}{2\sigma^2}\sum_{i=1}^n x_i^2 + \frac{\mu}{\sigma^2}\sum_{i=1}^n x_i - \frac{n\mu^2}{2\sigma^2})}$$
$$h(x) = (2\pi)^{-n/2}\sigma^{-n}e^{(\frac{-1}{2\sigma^2}\sum_{i=1}^n x_i^2)}$$

$$g(\sum_{i=1}^n x_i, \mu) = e^{\frac{\mu}{\sigma^2}\sum_{i=1}^n x_i - \frac{n\mu^2}{2\sigma^2}}$$

Easy to see $\bar{x} = \Sigma_i x_i/n$ is sufficient.

7 Ancilliarity

1. Suppose $\{Z_i\}_{i=1}^n$ is identically and independent distributed random variables with cdf $F(x)$. $X_i = Z_i + \theta$. $R = max(X_i) - min(X_i) = max(Z_i + \theta) - min(Z_i + \theta) = max(Z_i) - min(Z_i)$. R is a function of Z_i, thus it does not depend on θ.

8 Completeness

$$f(x_1, x_2, ..., x_n|\mu) = (2\pi|\mu|)^{-\frac{n}{2}} exp(\frac{1}{2\mu^2}\Sigma_i(x_i-\mu)^2 = (2\pi|\mu|)^{-\frac{n}{2}} exp(-\frac{n}{2|\mu|}(\bar{x}-\mu)^2)exp(-\frac{1}{2|\mu|^2}s^2)$$

Easy to know $(\bar{x}, s^2)$ is a sufficient statistic for $N(\mu, \mu^2)$.
Let $T = (\bar{x}, s^2)$, $h(T) = \bar{x}^2 - \frac{n+1}{n}s^2$,

$$E(h(T)) = E(\bar{x})^2 + Var(\bar{x}) - \frac{n+1}{n}E(s^2) = \mu^2 + \frac{\mu^2}{n} - \frac{n+1}{n}\mu^2 = 0$$

but $h(T)$ is not trivially 0.

9 Regular Exponential Family

1. Suppose $Y \sim Poisson(\lambda)$, the PMF is

$$p(x|\lambda) = \frac{\lambda^x e^{-\lambda}}{x!}$$

$$= \frac{1}{x!}e^{xlog\lambda-\lambda}$$

It follows the form

$$f(x|\theta) = h(x)e^{\psi(\theta)T(X)-A(\theta)}$$

if we let T(x)=x and $\psi(\theta) = log(\theta)$ and $A(\theta) = \theta$ and $h(x) = \frac{1}{x!}$

10 Regular Exponential Family

$$B(\eta) = log\int_x h(x)e^{\eta T(X)}dx$$

$$\frac{\delta}{\delta\eta_i}B(\eta) = \frac{\int_x T_i(x)h(x)e^{\eta T(X)}dx}{\int_x h(x)e^{\eta T(X)}dx} = E_i[T_i(X)]$$

$$\frac{\delta^2}{\delta\eta_i\eta_j}B(\eta) = \frac{\int_x T_i(x)T_j(x)h(x)e^{\eta T(X)}dx}{\int_x T_i(x)h(x)e^{\eta T(X)}dx} - \frac{(\frac{\delta}{\delta\eta_i})(\frac{\delta}{\delta\eta_j})}{(\int_x T_i(x)h(x)e^{\eta T(X)}dx)^2}$$

$$= E_\eta[T_i(X)T_j(X)] - E_\eta[T_i(X)]E_\eta[T_j(X)] = Cov_\eta[T_i(X), T_j(X)]$$

11 Delta Method

$X_i \sim Bernoulli(p)$,

$$n\bar{X} = X_1 + X_2 + ... + X_n \sim Binomial(n, p)$$

$$E(\bar{X}) = p$$

$$Var(\bar{X}) = \frac{p(1-p)}{n}$$

From Delta Method, we know

$$Var(\bar{X}(1-\bar{X})) = (1-2\hat{p})^2 Var(\bar{X}) = (1-2\hat{p})^2 \frac{p(1-p)}{n} = \frac{(1-2p)^2(1-p)p}{n^2}$$

Thus,

$$\sqrt{n}(\hat{p}(1-\hat{p}) - p(1-p)) \sim N(0, (1-2p)^2(1-p)p)$$

12 Joint Entropy

12.1

$H(X,Y) = \Sigma_x \Sigma_y P(x,y) log_2(P(x,y)) = P(0,0) log_2(P(0,0)) + P(1,0) log_2(P(1,0)) + P(2,0) log_2(P(2,0)) + P(0,1) log_2(P(0,1)) + P(1,1) log_2(P(1,1)) + P(2,1) log_2(P(2,1)) = 2 * \frac{1}{4} log_2 \frac{1}{4} + 2 * \frac{1}{6} log_2 \frac{1}{6} + 2 * \frac{1}{12} log_2 \frac{1}{12} = -2.45$

12.2

Marginal Distribution:
$P(X=1) = P(X=0, Y=0) + P(X=0, Y=1) = \frac{1}{3}$
$P(X=1) = P(X=1, Y=0) + P(X=1, Y=1) = \frac{1}{3}$
$P(X=1) = P(X=2, Y=0) + P(X=2, Y=1) = \frac{1}{3}$
Conditional Entropy:
$P(Y=0|X=0) = P(X=0, Y=0)/P(X=0) = 3/4$
$P(Y=0|X=1) = P(X=1, Y=0)/P(X=0) = 1/4$
$P(Y=0|X=2) = P(X=2, Y=0)/P(X=0) = 1/2$
$P(Y=1|X=0) = P(X=0, Y=1)/P(X=1) = 1/4$
$P(Y=1|X=1) = P(X=1, Y=1)/P(X=1) = 3/4$
$P(Y=1|X=2) = P(X=2, Y=1)/P(X=1) = 1/2$
$H(Y|X=0) = 3/4 * log_2(3/4) + 1/4 * log_2(1/4) = -0.81$
$H(Y|X=1) = 1/4 * log_2(1/4) + 3/4 * log_2(3/4) = -0.81$
$H(Y|X=2) = log_2(1/2) = -1$
$H(Y|X) = 1/3 * (-0.81 - 0.81 - 1) = -0.87$
$H(Y|X) = \Sigma_{y|x} P(y|x) log_2(P(y|x)) = 2 * \frac{3}{4} log_2 \frac{3}{4} + 2 * \frac{1}{4} log_2 \frac{1}{4} + 2 * \frac{1}{2} log_2 \frac{1}{2} = -2.62$

12.3

$H(X) = \Sigma_{y|x} P(y|x) log_2(P(y|x)) = log_2 1/3 = -1.58$
$H(X,Y) - H(X) = -2.45 - 1.58 = -0.87 = H(Y|X)$

13 Differential Entropy

$Differential\ Entropy = -\int_{-\infty}^{+\infty} N(x|\mu,\Sigma)\ln(N(x|\mu,\Sigma))dx = -E[\ln(N(x|\mu,\Sigma))] = -E[\ln((2\pi)^{-\frac{D}{2}}|\Sigma|^{-\frac{1}{2}}e^{-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)})] = \frac{D}{2}\ln(2\pi)+\frac{1}{2}\ln|\Sigma|+\frac{1}{2}E[(x-\mu)^T\Sigma^{-1}(x-\mu)]$

$E[(x-\mu)^T\Sigma^{-1}(x-\mu)] = E(tr(((x-\mu)^T\Sigma^{-1}(x-\mu))] = E(tr((\Sigma^{-1}(x-\mu)(x-\mu)^T)] = tr(E[\Sigma^{-1}(x-\mu)(x-\mu)^T]) = tr(\Sigma^{-1}E[(x-\mu)(x-\mu)^T]) = tr(\Sigma^{-1}\Sigma) = tr(I) = D$

$Differential\ Entropy = \frac{D}{2}\ln(2\pi) + \frac{1}{2}\ln|\Sigma| + D$
where D is the number of dimension for multivariate normal variable.