CSE 6363: Machine Learning
Problem Set 1 Solutions

Adhiraj Sen (1002264465)

Fall 2025

1 Problem 1: Maximum Likelihood Estimation for Gaussian Distribution

Given a Gaussian distribution $p(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$ and N i.i.d. observations $\{x_1, x_2, \ldots, x_N\}$, we derive the MLE estimates for μ and σ^2.

Likelihood Function:

$$L(\mu, \sigma^2) = \prod_{n=1}^{N} p(x_n|\mu, \sigma^2) = \prod_{n=1}^{N} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x_n-\mu)^2}{2\sigma^2}\right) \tag{1}$$

Log-likelihood:

$$\ln L(\mu, \sigma^2) = -\frac{N}{2}\ln(2\pi) - \frac{N}{2}\ln(\sigma^2) - \frac{1}{2\sigma^2}\sum_{n=1}^{N}(x_n-\mu)^2 \tag{2}$$

MLE for μ: Taking the derivative with respect to μ and setting to zero:

$$\frac{\partial \ln L}{\partial \mu} = \frac{1}{\sigma^2}\sum_{n=1}^{N}(x_n-\mu) = 0 \tag{3}$$

$$\sum_{n=1}^{N} x_n - N\mu = 0 \tag{4}$$

$$\mu_{ML} = \frac{1}{N}\sum_{n=1}^{N} x_n \tag{5}$$

MLE for σ^2: Taking the derivative with respect to σ^2 and setting to zero:

$$\frac{\partial \ln L}{\partial \sigma^2} = -\frac{N}{2\sigma^2} + \frac{1}{2(\sigma^2)^2}\sum_{n=1}^{N}(x_n-\mu)^2 = 0 \tag{6}$$

$$\frac{N}{\sigma^2} = \frac{1}{(\sigma^2)^2}\sum_{n=1}^{N}(x_n-\mu)^2 \tag{7}$$

$$\sigma^2_{ML} = \frac{1}{N}\sum_{n=1}^{N}(x_n-\mu_{ML})^2 \tag{8}$$

2 Problem 2: Least Squares Linear Regression

Given the linear model $y(x, w) = w_0 + w_1 x$ and error function $E(w) = \frac{1}{2}\sum_{n=1}^{N}\{t_n - y(x_n, w)\}^2$.

Expanding the error function:

$$E(w) = \frac{1}{2}\sum_{n=1}^{N}(t_n - w_0 - w_1 x_n)^2 \tag{9}$$

Taking partial derivatives and setting to zero:

$$\frac{\partial E}{\partial w_0} = -\sum_{n=1}^{N}(t_n - w_0 - w_1 x_n) = 0 \tag{10}$$

$$\frac{\partial E}{\partial w_1} = -\sum_{n=1}^{N} x_n(t_n - w_0 - w_1 x_n) = 0 \tag{11}$$

Rearranging:

$$N w_0 + w_1 \sum_{n=1}^{N} x_n = \sum_{n=1}^{N} t_n \tag{12}$$

$$w_0 \sum_{n=1}^{N} x_n + w_1 \sum_{n=1}^{N} x_n^2 = \sum_{n=1}^{N} x_n t_n \tag{13}$$

In matrix form: $\sum_{j=0}^{1} A_{ij} w_j = T_i$ where:

$$A_{00} = \sum_{n=1}^{N}(x_n)^0 = N, \quad A_{01} = \sum_{n=1}^{N} x_n \tag{14}$$

$$A_{10} = \sum_{n=1}^{N} x_n, \quad A_{11} = \sum_{n=1}^{N} x_n^2 \tag{15}$$

$$T_0 = \sum_{n=1}^{N} t_n, \quad T_1 = \sum_{n=1}^{N} x_n t_n \tag{16}$$

In general: $A_{ij} = \sum_{n=1}^{N}(x_n)^{i+j}$ and $T_i = \sum_{n=1}^{N}(x_n)^i t_n$.

3 Problem 3: Cross-entropy vs. Least Squares Derivatives

For binary classification with sigmoid activation $\sigma(a) = \frac{1}{1+e^{-a}}$ and $a = w^T x$:

Cross-entropy loss:

$$E_{CE} = -\sum_{n=1}^{N}[t_n \ln y_n + (1 - t_n)\ln(1 - y_n)] \tag{17}$$

where $y_n = \sigma(a_n)$ and $t_n \in \{0, 1\}$.

Derivative:

$$\frac{\partial E_{CE}}{\partial w_i} = -\sum_{n=1}^{N} \left[\frac{t_n}{y_n} - \frac{1-t_n}{1-y_n} \right] \frac{\partial y_n}{\partial w_i} \tag{18}$$

$$= -\sum_{n=1}^{N} \left[\frac{t_n}{y_n} - \frac{1-t_n}{1-y_n} \right] y_n(1-y_n)x_{ni} \tag{19}$$

$$= -\sum_{n=1}^{N} [t_n(1-y_n) - (1-t_n)y_n]x_{ni} \tag{20}$$

$$= \sum_{n=1}^{N} (y_n - t_n)x_{ni} \tag{21}$$

Least squares with sigmoid:

$$E_{LS} = \frac{1}{2}\sum_{n=1}^{N}(y_n - t_n)^2 \tag{22}$$

Derivative:

$$\frac{\partial E_{LS}}{\partial w_i} = \sum_{n=1}^{N}(y_n - t_n)\frac{\partial y_n}{\partial w_i} \tag{23}$$

$$= \sum_{n=1}^{N}(y_n - t_n)y_n(1-y_n)x_{ni} \tag{24}$$

The derivatives are the same when $y_n(1-y_n) = 1$, which occurs when the sigmoid is in its linear region or when we consider the canonical link function property of the logistic regression.

4 Problem 4: Categorical Distribution Likelihood

For a categorical distribution with K categories and parameters $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_K)$ where $\sum_{k=1}^{K} \pi_k = 1$:

Multinomial PMF for single observation:

$$p(\mathbf{x}|\boldsymbol{\pi}) = \prod_{k=1}^{K} \pi_k^{x_k} \tag{25}$$

where $\mathbf{x} = (x_1, \ldots, x_K)$ with $x_k \in \{0, 1\}$ and $\sum_{k=1}^{K} x_k = 1$.

Likelihood for N observations:

$$L(\boldsymbol{\pi}) = \prod_{n=1}^{N}\prod_{k=1}^{K} \pi_k^{x_{nk}} \tag{26}$$

$$= \prod_{k=1}^{K} \pi_k^{\sum_{n=1}^{N} x_{nk}} \tag{27}$$

$$= \prod_{k=1}^{K} \pi_k^{N_k} \tag{28}$$

where $N_k = \sum_{n=1}^{N} x_{nk}$ is the count of observations in category k.

5 Problem 5: Sigmoid Inverse is Logit

Given $\sigma(x) = \frac{1}{1+e^{-x}}$, we want to show that its inverse is $\text{logit}(x) = \ln\left(\frac{x}{1-x}\right)$.

Let $y = \sigma(x) = \frac{1}{1+e^{-x}}$. To find the inverse, solve for x in terms of y:

$$y = \frac{1}{1 + e^{-x}} \tag{29}$$

$$y(1 + e^{-x}) = 1 \tag{30}$$

$$y + ye^{-x} = 1 \tag{31}$$

$$ye^{-x} = 1 - y \tag{32}$$

$$e^{-x} = \frac{1-y}{y} \tag{33}$$

$$-x = \ln\left(\frac{1-y}{y}\right) \tag{34}$$

$$x = \ln\left(\frac{y}{1-y}\right) \tag{35}$$

Therefore, $\sigma^{-1}(y) = \ln\left(\frac{y}{1-y}\right) = \text{logit}(y)$.

Verification:

$$\sigma(\text{logit}(y)) = \sigma\left(\ln\left(\frac{y}{1-y}\right)\right) = \frac{1}{1 + e^{-\ln(y/(1-y))}} \tag{36}$$

$$= \frac{1}{1 + (1-y)/y} = \frac{1}{1/y} = y \quad \checkmark \tag{37}$$