CS280 Fall 2019 Assignment 1
Part A

Basic Neural Networks

Due on October 10, 2019, 23:59 UTC+8

Name:Zhang Chengrui

Student ID:2019233183

1. Hessian in Logistic Regression (10 points)

Let $\sigma(a) = \frac{1}{1+e^{-a}}$ be an activation function, the loss function of LR is

$$f(\mathbf{w}) = -\sum_{i=1}^{n}[y_i \log(\mu_i) + (1 - y_i)\log(1 - \mu_i)],$$

where $\mu_i = \sigma(\mathbf{w}^\intercal \mathbf{x}_i)$.

- Show that the Hessian of f can be written as $H = X^\intercal S X$, where $S = diag(\mu_1(1 - \mu_1), \cdots, \mu_n(1 - \mu_n))$ and $X = [X_1, \cdots, X_n]^\intercal$

Solution:

2. Linear Regression (5 points)

Linear regression has the form

$$f(x) = E[y|x] = b + \mathbf{w}^\intercal \mathbf{x}.$$

- It is possible to solve for $\mathbf{w}$ and b separately. Show that

$$b = \frac{1}{n}\sum_{i=1}^{n} y_i - \frac{1}{n}\sum_{i=1}^{n} \mathbf{x}_i^\intercal \mathbf{w} = \bar{y} - \bar{\mathbf{x}}^\intercal \mathbf{w}$$

Solution:

3. Gradient descent for fitting GMM (15 points)

Consider the Gaussian mixture model

$$p(\mathbf{x}|\theta) = \sum_k \pi_{k=1}^K \mathcal{N}(\mathbf{x}|\mu_k, \Sigma_k)$$

Define the log likelihood as

$$l(\theta) = \sum_{n=1}^N \log p(\mathbf{x}_n|\theta)$$

Denote the posterior responsibility that cluster k has for datapoint n as follows:

$$r_{nk} := p(z_n = k|\mathbf{x}_n, \theta) = \frac{\pi_k \mathcal{N}(\mathbf{x}_n|\mu_k, \Sigma_k)}{\sum_{k'} \pi_{k'} \mathcal{N}(\mathbf{x}_n|\mu_{k'}, \Sigma_k k')}$$

- Show that the gradient of the log-likelihood wrt μ_k is

$$\frac{d}{d\mu_k} l(\theta) = \sum_n r_{nk} \Sigma_k^{-1} (\mathbf{x}_n - \mu_k)$$

- Derive the gradient of the log-likelihood wrt π_k without considering any constraint on π_k. (bonus: with constraint $\sum_k \pi_k = 1$.)
- Derive the gradient of the log-likelihood wrt Σ_k without considering any constraint on Σ_k. (bonus: with constraint Σ_k be a symmetric positive definite matrix.)

Solution:

References