

Problem Set 2

1 Regularized Normal Equation for Linear Regression

Given a data set $\{x^{(i)}, y^{(i)}\}_{i=1, \dots, m}$ with $x^{(i)} \in \mathbb{R}^n$ and $y^{(i)} \in \mathbb{R}$, the general form of regularized linear regression is as follows

$$\min_{\theta} \frac{1}{2m} \left[\sum_{i=1}^m (h_{\theta}(x^{(i)}) - y^{(i)})^2 + \lambda \sum_{j=1}^n \theta_j^2 \right] \quad (1)$$

Derive the normal equation.

2 Gaussian Discriminant Analysis Model

Given m training data $\{x^{(i)}, y^{(i)}\}_{i=1, \dots, m}$, assume that $y \sim \text{Bernoulli}(\psi)$, $x | y = 0 \sim \mathcal{N}(\mu_0, \Sigma)$, $x | y = 1 \sim \mathcal{N}(\mu_1, \Sigma)$. Hence, we have

- $p(y) = \psi^y (1 - \psi)^{1-y}$
- $p(x | y = 0) = \frac{1}{(2\pi)^{n/2} |\Sigma|^{1/2}} \exp \left(-\frac{1}{2} (x - \mu_0)^T \Sigma^{-1} (x - \mu_0) \right)$
- $p(x | y = 1) = \frac{1}{(2\pi)^{n/2} |\Sigma|^{1/2}} \exp \left(-\frac{1}{2} (x - \mu_1)^T \Sigma^{-1} (x - \mu_1) \right)$

The log-likelihood function is

$$\begin{aligned} \ell(\psi, \mu_0, \mu_1, \Sigma) &= \log \prod_{i=1}^m p(x^{(i)}, y^{(i)}; \psi, \mu_0, \mu_1, \Sigma) \\ &= \log \prod_{i=1}^m p(x^{(i)} | y^{(i)}; \psi, \mu_0, \mu_1, \Sigma) p(y^{(i)}; \psi) \end{aligned}$$

Solve ψ , μ_0 , μ_1 and Σ by maximizing $\ell(\psi, \mu_0, \mu_1, \Sigma)$.

Hint: $\nabla_X \text{tr}(AX^{-1}B) = -(X^{-1}BAX^{-1})^T$, $\nabla_A |A| = |A|(A^{-1})^T$

3 MLE for Naive Bayes

Consider the following definition of **MLE problem for multinomials**. The input to the problem is a finite set \mathcal{Y} , and a weight $c_y \geq 0$ for each $y \in \mathcal{Y}$.

The output from the problem is the distribution p^* that solves the following maximization problem.

$$p^* = \arg \max_{p \in \mathcal{P}_{\mathcal{Y}}} \sum_{y \in \mathcal{Y}} c_y \log p_y$$

- (i) Prove that, the vector p^* has components

$$p_y^* = \frac{c_y}{N}$$

for $\forall y \in \mathcal{Y}$, where $N = \sum_{y \in \mathcal{Y}} c_y$. (Hint: Use the theory of Lagrange multiplier)

- (ii) Using the above consequence, prove that, the maximum-likelihood estimates for Naive Bayes model are as follows

$$p(y) = \frac{\sum_{i=1}^m \mathbf{1}(y^{(i)} = y)}{m}$$

and

$$p_j(x \mid y) = \frac{\sum_{i=1}^m \mathbf{1}(y^{(i)} = y \wedge x_j^{(i)} = x)}{\sum_{i=1}^m \mathbf{1}(y^{(i)} = y)}$$