Problem Set 5*

Tasha Pais†

Problem 1

We aim to show that for any probability distribution q,

$$R(hr; q) \geq R(h^*)$$

that is, the risk of the randomized classifier hr defined above is at least as high as the Bayes risk.

Proof:
Given:

$$R(hr; q) = \int_x \sum_{y=1}^{L} \sum_{y'=1}^{L} L_{0/1}(y', y) q(y'|x) p(x, y)\, dx$$

We need to show:

$$R(hr; q|x) \geq R(h^*|x) \text{ for any } x$$

By definition, the conditional risk $R(hr; q|x)$ is:

$$R(hr; q|x) = \sum_{y=1}^{L} \sum_{y'=1}^{L} L_{0/1}(y', y) q(y'|x) p(y|x)$$

The Bayes classifier $h^*(x)$ minimizes the risk, so for the Bayes risk $R(h^*|x)$, we have:

$$R(h^*|x) = \min_y p(y|x)$$

For the 0/1 loss, we know that:

$$L_{0/1}(y', y) = \begin{cases} 0 & \text{if } y' = y \\ 1 & \text{if } y' \neq y \end{cases}$$

Thus, the risk for the randomized classifier can be rewritten as:

$$R(hr; q|x) = 1 - \sum_{y=1}^{L} q(y|x) p(y|x)$$

*Due: December 13, 2023, Student(s) worked with: None

†NetID: tdp74, Email: tdp74@rutgers.edu

Since $\sum_{y=1}^{L} q(y|x) = 1$ and $p(y|x) \leq 1$, we have:

$$R(hr; q|x) \geq 1 - \max_y p(y|x)$$

But $\max_y p(y|x)$ is exactly the risk $R(h^*|x)$, therefore:

$$R(hr; q|x) \geq R(h^*|x)$$

Hence, it is shown that the risk of the randomized classifier hr is at least as high as the Bayes risk for any x.

Problem 2

We need to show that the posterior $p(1|x)$ resulting from the equal-covariance Gaussian generative model is equivalent to the logistic regression model, i.e.,

$$p(1|x) = \frac{1}{1 + \exp(-w^\top x - b)}$$

for some $w \in \mathbb{R}^d$ and $b \in \mathbb{R}$.

Proof:

Given the Gaussian generative model, the likelihood for a class y is:

$$p(x|y) = \frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(x - \mu_y)^\top \Sigma^{-1}(x - \mu_y)\right)$$

Using Bayes' rule, the posterior is:

$$p(y|x) = \frac{p(x|y)p(y)}{p(x)}$$

For binary classification ($L = 2$), focusing on $p(1|x)$:

$$p(1|x) = \frac{p(x|1)p(1)}{p(x|1)p(1) + p(x|2)p(2)}$$

Substituting the Gaussian likelihoods and simplifying:

$$p(1|x) = \frac{1}{1 + \frac{p(x|2)p(2)}{p(x|1)p(1)}}$$

Expressing the ratio inside the exponential:

$$p(1|x) = \frac{1}{1 + \exp\left(\log \frac{p(x|2)p(2)}{p(x|1)p(1)}\right)}$$

Expanding the logarithm and using the Gaussian likelihoods:

$$p(1|x) = \frac{1}{1 + \exp\left(-\frac{1}{2}(x - \mu_1)^\top \Sigma^{-1}(x - \mu_1) + \frac{1}{2}(x - \mu_2)^\top \Sigma^{-1}(x - \mu_2) + \log \frac{p(2)}{p(1)}\right)}$$

This can be rewritten as:

$$p(1|x) = \frac{1}{1 + \exp(-w^\top x - b)}$$

where

$$w = \Sigma^{-1}(\mu_1 - \mu_2)$$

and

$$b = \frac{1}{2}(\mu_2^\top \Sigma^{-1} \mu_2 - \mu_1^\top \Sigma^{-1} \mu_1) + \log \frac{p(2)}{p(1)}$$

Hence, we have shown that $p(1|x)$ from the Gaussian generative model has the same form as the logistic regression model.

Problem 3

The question is whether logistic regression and linear discriminant analysis (LDA) based on the equal-covariance Gaussian model will produce the same classifier when learned from a given training set.

Answer:

Although both logistic regression and the equal-covariance Gaussian model-based LDA lead to classifiers with similar forms of the posterior $p(y|x)$, they generally do not produce the same classifier when learned from the same training set. The key differences arise from their learning approaches and assumptions.

Differences in Learning:

- **Logistic Regression:** It estimates the parameters directly by maximizing the likelihood of the observed data. It does this without making strong assumptions about the distribution of the predictor variables.

- **LDA:** It assumes that the predictor variables are normally distributed and that different classes share the same covariance matrix. LDA estimates the parameters (means and shared covariance matrix) based on these assumptions.

Implications:

1. **Parameter Estimation:** Since logistic regression and LDA use different methods for parameter estimation, the resulting classifiers will differ unless the data perfectly fits the assumptions made by LDA.

2. **Assumptions:** The equal-covariance Gaussian assumption of LDA is quite restrictive compared to the flexibility of logistic regression. If the true data distribution deviates from these assumptions (e.g., non-Gaussian features or unequal covariances), LDA's performance may be suboptimal compared to logistic regression.

3. **Data Sensitivity:** Logistic regression is more robust to deviations from Gaussian distributions and is less sensitive to outliers compared to LDA.

Conclusion:

While the two models have the same form of the posterior $p(y|x)$, the differences in assumptions and learning methods lead to different classifiers when trained on the same dataset, except in special cases where the data perfectly aligns with LDA's assumptions.