Formulas from 'Generative Linear and Quadratic Classifiers'

1 Bayesian Classification

Optimal Bayes Decision Rule: Assign the class with highest posterior probability

$$c_t^* = \arg\max_c P(C_t = c|X_t = x_t) \tag{1}$$

Bayes Theorem for Classification:

$$P(C_t = c|X_t = x_t) = \frac{f_{X,C}(x_t, c)}{\sum_{c'} f_{X,C}(x_t, c')} \tag{2}$$

Joint Density Expression:

$$f_{X_t,C_t}(x_t, c) = f_{X|C}(x_t|c)P(C = c) \tag{3}$$

2 Gaussian Classifiers

2.1 Gaussian Class Conditional Distribution:

$$(X|C = c) \sim N(\mu_c, \Sigma_c) \tag{4}$$

2.2 Maximum Likelihood Estimates for Gaussian Model:

A possible way to estimate the model parameters is to maximize the data log-likelihood:
Log-Likelihood Function:

$$\ell(\theta) = \sum_{c=1}^{k} \sum_{i|c_i=c} \log N(x_i|\mu_c, \Sigma_c) \tag{5}$$

2.3 MLE solution

$$\mu_c^* = \frac{1}{N_c} \sum_{i|c_i=c} x_i \tag{6}$$

$$\Sigma_c^* = \frac{1}{N_c} \sum_{i|c_i=c} (x_i - \mu_c^*)(x_i - \mu_c^*)^T \tag{7}$$

3 Log-Likelihood Ratio for Classification

Posterior Ratio:

$$\log r(x_t) = \log \frac{P(C = h_1|x_t)}{P(C = h_0|x_t)} \tag{8}$$

Decomposed Posterior Ratio:

$$\log r(x_t) = \log \frac{f_{X|C}(x_t|h_1)}{f_{X|C}(x_t|h_0)} + \log \frac{P(C = h_1)}{P(C = h_0)} \tag{9}$$

The second term represents the prior (log)-odds. For a binary problem, we have

$$P(C = h_1) = \pi, \quad P(C = h_0) = 1 - P(C = h_1) = 1 - \pi$$

thus

$$\log r(\mathbf{x}_t) = \log \frac{f_{X|C}(\mathbf{x}_t|h_1)}{f_{X|C}(\mathbf{x}_t|h_0)} + \log \frac{\pi}{1-\pi}$$

The optimal decision is based on the comparison

$$\log r(\mathbf{x}_t) \gtrless 0$$

Decision Boundary for Gaussian Classifier:

$$\operatorname{llr}(x) = x^T A x + x^T b + c \tag{10}$$

4 Naive Bayes Assumption

Naive Bayes Factorization:

$$f_{X|C}(x|c) = \prod_{j=1}^{D} f_{X[j]|C}(x[j]|c) \tag{11}$$

Naive Bayes Gaussian Model:

$$f_{X[j]|C}(x[j]|c) = N(x[j]|\mu_{c,[j]}, \sigma^2_{c,[j]}) \tag{12}$$

The **naive Bayes Gaussian** classifier corresponds to a Multivariate Gaussian classifier with **diagonal** covariance matrices (this does not hold for a generic density!). The decision boundary is still quadratic. The ML solution is

$$\mu^*_{c,[j]} = \frac{1}{N_c} \sum_{i|c_i=c} x_{i,[j]}, \quad \sigma^2_{c,[j]} = \frac{1}{N_c} \sum_{i|c_i=c} \left(x_{i,[j]} - \mu_{c,[j]}\right)^2$$

5 Tied Gaussian Assumption

Another common Gaussian model assumes that the covariance matrices of the different classes are tied

The tied covariance model assumes that

$$f_{X|C}(\mathbf{x} \mid c) = \mathcal{N}(\mathbf{x} \mid \boldsymbol{\mu}_c, \boldsymbol{\Sigma})$$

i.e., each class has its own mean $\boldsymbol{\mu}_c$, but the covariance matrix is the same for all classes.

Again, we can estimate the parameters using the ML framework. In this case, the log-likelihood does not factorize over classes.

The ML solution is:

$$\boldsymbol{\mu}^*_c = \frac{1}{N_c} \sum_{i|c_i=c} \mathbf{x}_i, \quad \boldsymbol{\Sigma}^* = \frac{1}{N} \sum_c \sum_{i|c_i=c} (\mathbf{x}_i - \boldsymbol{\mu}_c)(\mathbf{x}_i - \boldsymbol{\mu}_c)^T$$

where N is the number of samples:

$$N = \sum_{c=1}^{k} N_c$$

The decision boundary is now linear:

$$\begin{aligned} llr(\mathbf{x}) &= \log \frac{f_{X|C}(\mathbf{x} \mid h_1)}{f_{X|C}(\mathbf{x} \mid h_0)} \\ &= \mathbf{x}^T \mathbf{b} + c \end{aligned}$$

6 Tied Gaussian and LDA

They are equivalent in the sense that they result in the same decision boundaries under the assumption of Gaussian-distributed data with a shared covariance matrix.

However, the key difference is:

- **Tied Gaussian classification** is a generative approach that models the likelihoods directly.
- **LDA** is derived from an optimization perspective, finding the best projection for class separation while also leading to linear decision boundaries.

If your goal is simply classification, a **tied Gaussian model** is functionally the same as **LDA** in that both use a common covariance matrix assumption and result in linear decision boundaries. However, their interpretations and derivations are different.

7 Categorical and Multinomial Models

Categorical Distribution Model:

$$P(X_t = x_t | C_t = c) = \pi_{c,x_t} \tag{13}$$

Maximum Likelihood Estimate for Categorical Model:

$$\ell_c(\pi_c) = \sum_{i|c_i=c} \log \pi_{c,x_i}$$

The ML solution is given by:

$$\pi^*_{c,i} = \frac{N_{c,i}}{N_c} \tag{14}$$

Multinomial Distribution Model:

If we have more than one categorical attribute, we may model their joint probability as a categorical R.V. with values given by all possible combinations of the attributes.

In practice, the number of elements would quickly become intractable.

We can adopt again a naive Bayes approximation and assume that features are independent.

Log-Likelihood Ratio for Multinomial Model:

$$\ell_c(\pi_c) = \sum_{i|c_i=c} \sum_{j=1}^{m} x_{i,[j]} \log \pi_{c,j}$$

The ML solution is given by:

$$\pi^*_{c,j} = \frac{N_{c,j}}{N_c} \tag{15}$$