Lecture 4

Optimal Bayes Classifier

Decision Rule

- classifier assigns an unknown pattern $\boldsymbol{x}$ to class c_i if
  $r_i(\boldsymbol{x}) < r_j(\boldsymbol{x})$ for $j = 1, 2, \ldots, N; j \neq i$.

$$p(\boldsymbol{x}) - p(\boldsymbol{x}|c_i)P(c_i) < p(\boldsymbol{x}) - p(\boldsymbol{x}|c_j)P(c_j),$$

or equivalently

$$\boxed{p(\boldsymbol{x}|c_i)P(c_i) > p(\boldsymbol{x}|c_j)P(c_j)}$$

The *covariance matrix* of two random variables is the matrix of pairwise covariance calculations between each variable,

$$C = \begin{pmatrix} \mathrm{cov}(A, A) & \mathrm{cov}(A, B) \\ \mathrm{cov}(B, A) & \mathrm{cov}(B, B) \end{pmatrix}.$$

Symmetric

$$\mathrm{cov}(A, B) = E[(A-\bar{A})(B-\bar{B})] = E[AB - A\bar{B} - \bar{A}B + \bar{A}\bar{B}] = E[AB] - \bar{A}\bar{B} - \bar{A}\bar{B} + \bar{A}\bar{B} = E[AB] - E[A]E[B]$$

$$\mathrm{cov}(A, A) = E[(A-\bar{A})^2] = \mathrm{Var}(A)$$

- If the covariance matrix is identical. then

$$d_j(\mathbf{x}) = \ln P(\omega_j) + \mathbf{x}^T\mathbf{C}^{-1}\mathbf{m}_j - \frac{1}{2}\mathbf{m}_j^T\mathbf{C}^{-1}\mathbf{m}_j$$

- If all classes are equally likely and the covariance matrix is an identity matrix, then

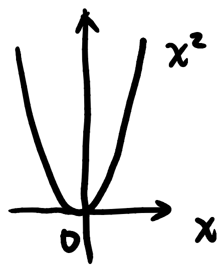
$$d_j(\mathbf{x}) = \mathbf{x}^T\mathbf{m}_j - \frac{1}{2}\mathbf{m}_j^T\mathbf{m}_j \qquad j = 1, 2, \ldots, W$$

- The same decision function for a minimum-distance classifier, which is optimal in the Bayes sense if
  - The pattern classes are Gaussian.
  - All covariance matrices are equal to identity matrix.
  - All classes are equally likely.

$$f(x) = x^2$$

$$f'(x) = 2x = 0 \quad \Rightarrow x = 0$$

$$f''(x) = (2x)' = 2 > 0 \quad \Rightarrow \quad f(0) = \min f(x)$$



Matrix calculus:

For a scalar α given by a quadratic form: $\alpha = \mathbf{x}^T \mathbf{A} \mathbf{x}$

$(1\times n \;\; n\times n \;\; n\times 1) \rightarrow 1\times 1$

where $\underbrace{\mathbf{x} \text{ is } \mathrm{n} \times 1}_{X^T \,:\, 1\times n}$, $\mathbf{A}$ *is* $\mathrm{n} \times \mathrm{n}$, *and* $\mathbf{A}$ *does not depend on* $\mathbf{x}$, *then*

$$\frac{\partial \alpha}{\partial \mathbf{x}} = \underbrace{\underbrace{\mathbf{x}^T}_{1\times n} \underbrace{(\mathbf{A} + \mathbf{A}^T)}_{n\times n}}_{1\times n}$$

$$X = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \quad A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix},$$

$$\alpha = \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} x_1 + \underbrace{3x_2} & 2x_1 + 4x_2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

$$= x_1 (x_1 + 3x_2) + x_2 (2x_1 + 4x_2)$$

$$= x_1^2 + 5x_1 x_2 + 4x_2^2$$

$$\frac{\partial \alpha}{\partial X} = \begin{bmatrix} \frac{\partial \alpha}{\partial x_1} & \frac{\partial \alpha}{\partial x_2} \end{bmatrix} = \begin{bmatrix} 2x_1 + 5x_2 & 5x_1 + 8x_2 \end{bmatrix}$$

$$A + A^T = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} + \begin{bmatrix} 1 & 3 \\ 2 & 4 \end{bmatrix} = \begin{bmatrix} 2 & 5 \\ 5 & 8 \end{bmatrix}$$

$$A + A^T = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} + \begin{bmatrix} 1 & 3 \\ 2 & 4 \end{bmatrix} = \begin{bmatrix} 2 & 5 \\ 5 & 8 \end{bmatrix}$$

$$x^T (A + A^T) = (x_1 \quad x_2] \begin{bmatrix} 2 & 5 \\ 5 & 8 \end{bmatrix} = (2x_1 + 5x_2 \quad 5x_1 + 8x_2)$$

http://scikit-learn.org › stable › modules › naive_bayes

1.9. Naive Bayes — scikit-learn 1.1.2 documentation

Naive Bayes learners and classifiers can be extremely fast compared to more sophisticated methods. The decoupling of the class conditional feature distributions ...

Sklearn.naive_bayes

Gaussian Naive Bayes (GaussianNB). Can perform ...

naive_bayes.MultinomialNB

Naive Bayes classifier for multinomial models. The ...

naive_bayes.BernoulliNB

Naive Bayes classifier for multivariate Bernoulli models ...

naive_bayes.CategoricalNB

Naive Bayes classifier for categorical features. The ...