Bayesian Classification

assign sample to most probable class

general definitions:

patterns (features): $\bar{x} = [x_1, x_2, \ldots, x_N]$

measures on shape or object

classes: C classes $C = \{\omega_1, \omega_2, \ldots, \omega_C\}$

priors: $p(\omega_i)$: probability of drawing a feature vector from class ω_i

training set $S = \{\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_M\}$

M feature vectors whose classes are known

Bayes Decision Rule

if $p(\omega_j|\bar{x}) > p(\omega_k|\bar{x}) \quad \forall k \neq j$

then assign $\bar{x}$ to ω_j

where $p(\omega_j|\bar{x})$ probability of class ω_j given $\bar{x}$

Use Bayes theorem:

$$p(\omega_j|\bar{x}) = \frac{p(\bar{x}|\omega_j)\,p(\omega_j)}{p(\bar{x})}$$

---

in the rule:

if $p(\bar{x}|\omega_j)p(\omega_j) > p(\bar{x}|\omega_k)P(\omega_k) \quad \forall k \neq j$

then assign $\bar{x}$ to ω_j

don't need division by $p(\bar{x})$ since it's in both sides

14|12

$p(\bar{x} \mid w_j)$: probability of vector $\bar{x}$
given class w_j

$p(w_j)$: priors are determined by frequency
e.g., 'e' is most probable letter
in english 12.7%

$p(\bar{x})$: $p(\bar{x}) = \sum_j p(\bar{x} \mid w_j) p(w_j)$

Multi-variate normal distribution

mvnrnd(mu, Sigma, n)

e.g., [0;0], eye(2), 1000

$$p(\bar{x}) = \frac{1}{2\pi^{\frac{N}{2}} |C|^{1/2}} e^{-\frac{1}{2}(\bar{x}-\bar{\mu})^T C^{-1} (\bar{x}-\bar{\mu})}$$

$\bar{\mu}$ is distribution mean C is covariance matrix

---

mvnpdf(x, mu, Sigma)

Bayesian classifiers : multivariate normal distribution

if $p(\bar{x}|w_j)\,p(w_j) > p(\bar{x}|w_k)\,p(w_k) \quad \forall k \neq j$

then assign $\bar{x}$ to w_j

taking log:

if $\log(p(\bar{x}|w_j)) + \log(p(w_j)) > \log(p(\bar{x}|w_k)) + \log(p(w_j))$

then assign $\bar{x}$ to w_j

$$\log(p(\bar{x}|w_j)) + \log(p(w_j))$$

$$= -\frac{1}{2}(\bar{x}-\bar{\mu})^T C_j^{-1}(\bar{x}-\bar{\mu}) - \frac{N}{2}\log 2\pi - \frac{1}{2}\log|C_j|$$
$$+ \log(p(w_j))$$

Simpler cases $\quad C_j = \sigma^2 I$

then $\quad = -\dfrac{\|\bar{x}-\bar{\mu}\|^2}{2\sigma^2} + \log(p(w_j))$

$$= \bar{w}_j\bar{x} + v_j \qquad \bar{w}_j = \frac{\bar{\mu}_j}{\sigma^2}$$

$$v_j = \frac{1}{2\sigma^2}\bar{\mu}_j^T\bar{\mu}_i + \log(p(w_j))$$

If $C_j = C \quad \forall j$

$$= \bar{\omega}_j^T \bar{x} + v_j$$

where $\bar{\omega}_j = C^{-1} \bar{\mu}_j$

$$v_j = -\frac{1}{2} \bar{\mu}_j^T C \bar{\mu}_j + \log(p(\omega_j))$$

Fisher linear discriminant from N-D to 1D

Given feature vectors, $\{\bar{x}_i\}$ $\quad \bar{x}_i \in \mathbb{R}^N$

project them onto an axis

so they are maximally separated.

$\bar{y} = \bar{w}^T \bar{x}$ $\qquad \bar{y}$ is new vector

criterion : $J(\bar{w}) = \dfrac{|\bar{m}_A - \bar{m}_B|^2}{s_A^2 + s_B^2}$ (2 classes)

$\bar{m}_A, \bar{m}_B$ are means ; s_A^2, s_B^2 sample variances

then: $\bar{w} = S_W^{-1} [\bar{M}_A - \bar{M}_B]$ original class means

$S_W = S_A + S_B$

Consider chars45.mat data (i.e., mask)

classes = 50

what is $P(\omega_k)$ for class k?

Use data in mask for this:

from ground_truth: get # A's
B's etc.

then $\frac{\#\text{ char } k}{\text{total}\#\text{ chars}} = P(\omega_k)$

e.g., ~~#~~ A's is 17
total # chars is 749 $\Big\}$ $P(\text{'A'}) = 2.27\%$

---

area

A's: $\mu_A = 172.0588$

$\sigma_A^2 = 154.5588$

$\mu_E = 228.5821$

$\sigma_E^2 = 89.1560$

---

perimeter

$\mu_A = 83.0588$

$\sigma_A^2 = 18.1838$

$\mu_E = 89.67$

$\sigma_E^2 = 10.133$