D PR

20.04.2022

slide 1.22.1

Performance measures for $\hat{H}$

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

If $A \cap B = \phi$ then

$$P(A \cup B) = P(A) + P(B)$$

Hence,

$$P_{TN} + P_{FP} = P(\hat{H} = H_0 \mid H = H_0) + P(\hat{H} = H_1 \mid H = H_0)$$

$$= P(\underbrace{\hat{H} = H_0 \text{ or } H_1}_{\text{this is always true due to binary detection}} \mid H = H_0)$$

$$= 1$$

$$P_{FN} + P_{TP} = 1 = P(\hat{H} = H_0 \text{ or } H_1 \mid H = H_1)$$

i.e.

$P(\hat{H} \mid H)$	$H = H_0$	$H = H_1$	Σ
$\hat{H} = H_0$	$P_{TN} = 1 - P_{FP}$	$P_{FN} = 1 - P_{TP}$	$\neq 1$
$\hat{H} = H_1$	P_{FP}	P_{TP}	$\neq 1$
Σ	1	1	no guarantee to be 1

Goal: Design $\hat{H}(\underline{x})$ such that

$\left.\begin{array}{l} *)\ \max\ P_{TP} \\ *)\ \min\ P_{FP} \end{array}\right\}$ impossible to achieve both of them at the same time. We can only compromise, due to 2D complex optimisation problem. Conflicting goals.

This can be extended to Multinary detection problem.

1.3 Pattern Recognition = ML.

1.25 No mathematical way to tell what is the best way for feature extraction. Domain knowledge is important here.

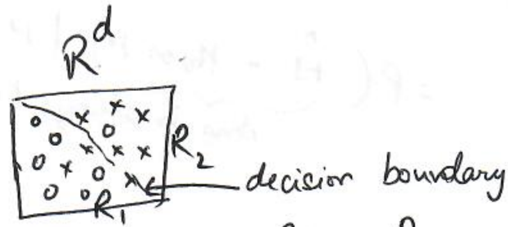
We need good generalized boundary so as to not overfit or underfit.

Q How can we reach to the generalized decision boundary

Q Will there be only 1 decision rule for 1 model for a given dataset?

1.30.1

Classification



- Divide feature space R^d into C non-overlapping decision regions $R_1 \ldots R_C$
- If the new feature vector $\underline{x} \in R_i$ then $\hat{\omega}(\underline{x}) = \omega_i$

decision rule

How to determine / design decision rule?

detection	vs	pattern recognition
from signal model $p(\underline{x} \mid H_0)$, $p(\underline{x} \mid H_1)$		learn from examples. → need a dataset/database

database of raw data
↓ feature extraction
database of feature vectors $\underline{x}_1, \underline{x}_2 \ldots \underline{x}_n$

Divide the whole data into 2 non-overlapping parts - training data, test data

When the Test & Trainging error rate is nearby, then the m/c is ready for deployment. If train ~~≈~~ test then overfitting.

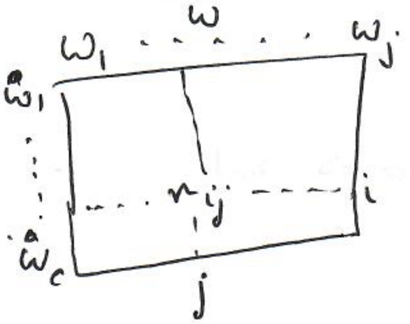
1.35.1 / 1.4 Confusion matrix

A detailed performance measure for ~~both~~ detector / classifier.

Confusion matrix ω

$[n_{ij}]_{ij}$ $C \times C$ matrix

$\hat{\omega}$



$n_{ij} = \#(\hat{\omega} = \omega_i, \omega = \omega_j) \in \mathbb{N}$

$\#$ no of occurrence

column sum $= n_{:j} = \sum_{i=1}^{c} n_{ij} = \#(\omega = \omega_j)$

row sum $= n_{i:} = \sum_{j=1}^{c} n_{ij} = \#(\hat{\omega} = \omega_i)$

matrix sum $= N = \sum_{j=1}^{c}\sum_{i=1}^{c} n_{ij}$ = total no of ~~xxx~~ samples.

Normalization of $n_{ij} \to$ probability.

(a) marginal prob: $P_j = P(\omega = \omega_j) = \cancel{\lim_{N\to\infty}} \dfrac{n_{:j}}{N}$

$\sum P_j = 1$

$P(\hat{\omega} = \omega_i) = \cancel{\lim_{N\to\infty}} \dfrac{n_{i:}}{N}$ $\sum P_i = 1$

(b) Joint prob: by matrix sum normalization

$P_{ij} = P(\hat{\omega} = \omega_i, \omega = \omega_j) = \dfrac{n_{ij}}{N}$ $\sum\sum P_{ij} = 1$

(c) Conditional. Probability :

column sum normalization : $\bar{P}_{ij} = P(\hat{\omega} = \omega_i \mid \omega = \omega_j)$

$$= \frac{n_{ij}}{n_{:j}} \triangleq TN, FP, FN, TP \text{ rate}$$

row sum naormalization $P(\hat{\omega} = \omega_j \mid \hat{\omega} = \omega_i)$

$$= \frac{n_{ij}}{n_{i:}}$$

1.38

$n_{11} \rightarrow$ not needed as not a requirement

$F_\beta \rightarrow$ weighted version of F_1

1.38.1

	ω: ω_1	ω_2
$\hat{\omega}$: ω_1	n_{11}	n_{12}
ω_2	n_{21}	n_{22}

precision $\frac{n_{22}}{n_{21} + n_{22}}$

$P_{FP} = \frac{n_{21}}{n_{11} + n_{21}}$ $\quad$ $P_{TP} = \frac{n_{22}}{n_{12} + n_{22}}$ = recall

Need to maximize F_1