Computing the Bayes error:

Quantifies the expected performance.

We saw that for a 0-1 loss case,

$$\text{error} = \int_{\mathbb{R}^n} \min\left(P_0 f_0(x), P_1 f_1(x)\right) dx.$$

This is a difficult integral to evaluate in general.
However can be computed for some special cases.

Neyman-pearson Criterion.

We may want to explicitly trade one type of error with another.

Eg: minimize Type-II error under the constraint that Type-I is < some threshold.

Type I error : False positive

Type II error : False negative

N-P : minimize false negatives under a fixed constraint for false positives.

Mathematically, Given any $\alpha \in (0,1)$

1. $P\left[h_{NP}(x) = 1 \mid x \in C\text{-}0\right] \leq \alpha$

2. $P\left[h_{NP}(x) = 0 \mid x \in C\text{-}1\right] \leq P\left[h(x) = 0 \mid x \in C_1\right]$

$\forall$ h such that $P\left[h(x)=1 \mid x \in C_0\right] \leq \alpha$.

<u>Claim:</u> $h_{NP}(x) = 1$ if $\dfrac{f_1(x)}{f_0(x)} > K$

$= 0$ otherwise

where K is such that

$$P\left[\frac{f_1(x)}{f_0(x)} \leq K \,\middle|\, x \in C\text{-}0\right] = 1-\alpha.$$

By, construction,

Proof: $P\left[h_{NP}(x)=1 \mid X \in C\text{-}0\right]$

$$= P\left[\frac{f_1(x)}{f_0(x)} > k \;\middle|\; X \in C\text{-}0\right]$$

$$= \alpha$$

Needs to be shown: Type-II error is < any other classifier satisfying this constraint on type-I error.

Let h be any classifier s.t

$$P\left[h(x)=1 \mid X \in C\text{-}0\right] \leq \alpha.$$

To complete the proof, to show

$$P\left[h_{NP}(x)=0 \mid X \in C\text{-}1\right] \leq P\left[h(x)=0 \mid X \in C\text{-}1\right]$$

or $P\left[h_{NP}(x)=1 \mid X \in C\text{-}1\right] \geq P\left[h(x)=1 \mid X \in C\text{-}1\right].$

Consider the integral,

$$I = \int_{\mathbb{R}^n} \left[h_{NP}(x) - h(x) \right] \left(f_1(x) - k f_0(x) \right) dx$$

$$= \int_{f_1 > k f_0} \left(h_{NP}(x) - h(x) \right) \left(f_1(x) - k f_0(x) \right) dx$$

$$+ \int_{f_1 \leq k f_0} \left(h_{NP}(x) - h(x) \right) \left(f_1(x) - k f_0(x) \right) dx$$

Lets first show that this integral is always non-negative.

When $f_1(x) > k f_0(x)$, we have

$$h_{NP}(x) - h(x) = 1 - h(x) \geq 0.$$

$$\Rightarrow \left(h_{NP}(x) - h(x) \right) \left(f_1(x) - k f_0(x) \right)$$

IIly, when $f_1(x) < k f_0(x)$.

$$h_{NP}(x) - h(x) = 0 - h(x) \leq 0$$

$$\Rightarrow \left(h_{NP}(x) - h(x) \right) \left(f_1(x) - k f_0(x) \right) \geq 0.$$

(24). $\Rightarrow I \geq 0.$

Thus, we have

$$\int_{R^n} \left(h_{NP}(x) - h(x) \right) \left(f_1(x) - k f_0(x) \right) dx \geq 0$$

$$\Rightarrow \int h_{NP}(x) f_1(x)\, dx - \int h(x) f_1(x)\, dx \geq$$

$$k \left[\int h_{NP}(x) f_0(x)\, dx - \int h(x) f_0(x)\, dx \right]$$

Since h_{NP} & $h \in \{0, 1\}$.

$$\int_{R^n} h_{NP}(x) f_1(x)\, dx = P\left[h_{NP}(x) = 1 \,\middle|\, x \in C_{-1} \right]$$

&

$$\int_{R^n} h(x) f_1(x)\, dx = P\left[h(x) = 1 \,\middle|\, x \in C_{-1} \right].$$

$\|^{ly}$ for integrals involving f_0.

We have already shown that,

$$\int h_{NP}(x) f_1(x)\, dx - \int h(x) f_1(x)\, dx \geq$$

$$K \left[\int h_{NP}(x) f_0(x)\, dx - \int h(x) f_0(x)\, dx \right].$$

Hence,

$$P\left[h_{NP}(x) = 1 \mid x \in C\text{-}1\right] - P\left[h(x) = 1 \mid x \in C\text{-}1\right]$$

$$\geq K \left[P\left[h_{NP}(x) = 1 \mid x \in C\text{-}0\right] - P\left[h(x) = 1 \mid x \in C\text{-}0\right] \right].$$

Hence we have,

$$P\left[h_{NP}(x) = 1 \mid x \in C\text{-}1\right] - P\left[h(x) = 1 \mid x \in C\text{-}1\right]$$

$$\geq K \left[P\left[h_{NP}(x) = 1 \mid x \in C\text{-}0\right] - P\left[h(x) = 1 \mid x \in C\text{-}0\right] \right].$$

But we have shown that the RHS is non-negative for all conditions.

Hence, $P\left[h_{NP}(x) = 1 \mid x \in C\text{-}1\right] - P\left[h(x) = 1 \mid x \in C\text{-}1\right] \geq 0$

Hence the proof.

Recall: Bayes classifier – General case.

$$h_B(x) = \alpha_i \quad \text{if}$$

$$R(\alpha_i | x) \leq R(\alpha_j | x) \ \forall j$$

where $R(\alpha_i | x) = \sum_{j=0}^{M-1} L(C_j, \alpha_i)\, q_j(x).$

Example: K-classes., K+1th class – reject the pattern.

$$L(i,j) = 0 \quad \text{if} \quad i=j, \ i,j = 1, \ldots, K$$
$$= \rho_m \quad \text{if} \quad i = 1, \ldots K, \ \& \ i \neq j$$
$$= \rho_r \quad \text{if} \quad i = K+1.$$

Goal: Derive the BC in terms of qs.

For, $\alpha_i = 1, \ldots, K$, we have $L(C_j, \alpha_i) = \rho_m$. if $\alpha_i \neq C_j$

Hence, $R(i | x) = \sum_{j \neq i} \rho_m q_j(x) = \rho_m \left[1 - q_i(x)\right]$

Also, $R(K+1 | x) = \sum_j \rho_r q_j(x) = \rho_r.$

Hence, $h_B(x) = i, \quad 1 \le i \le K$, if

$$p_m\left(1 - q_i(x)\right) \le p_m\left(1 - q_j(x)\right) \; \forall j$$

$$\& \quad p_m\left(1 - q_i(x)\right) \le p_r.$$

Thus, $h_B(x) = i, \quad 1 \le i \le K$ if

(i) $q_i(x) \ge q_j(x) \quad \forall j$

(ii) $q_i(x) \ge 1 - \frac{p_r}{p_m}$;

else $\underline{h_B(x) = K+1}$.

Finding the Bayes error.

Gives the expected performance.

Eg: for 0-1 loss f^n, ω

$$\text{error} = \int_{\mathbb{R}^n} \min\left(P_0 f_0(x), P_1 f_1(x)\right) dx.$$

Difficult to find in general cases.

Eg: 2-class problem, $X \in \mathbb{R}$, $f_i(x) \sim N(\mu_i, \sigma)$.
0-1 loss, $P_1 = P_0$.

We have seen that $h_B(x) = 0$ if
$$X < \frac{\mu_0 + \mu_1}{2}.$$

$$\therefore P(\text{error}) = R(h_B) = \frac{1}{2} \int_{-\infty}^{\frac{\mu_0+\mu_1}{2}} f_1(x)\, dx + \frac{1}{2} \int_{\frac{\mu_0+\mu_1}{2}}^{\infty} f_0(x)\, dx$$

$$= 0.5\, \phi\left(\frac{\mu_0 - \mu_1}{2\sigma}\right) + 0.5\left[1 - \phi\left(\frac{\mu_1 - \mu_0}{2\sigma}\right)\right].$$

$Z = \frac{X - \mu_i}{\sigma}$ converts the integrals into standard normal.

Thus, $\frac{|\mu_0 - \mu_1|}{\sigma}$ is called discriminability

Discuss NP here.

ROC - Curves.

In a Binary Class problem there are 4 - types of cases:

e_0 (Miss) | False negative.

e_1 False alarm | False positive.

$1-e_0$ Correct detection | True positive..

$1-e_1$ correct rejection | True negative.

For a 1-D feature space with 2-class consider $h(X)=0$ if $X<\tau$. For the previous case of Gaussians.

$$P[\text{error}] = 0.5\,\phi\left(\frac{\tau-\mu_1}{\sigma}\right) + 0.5\left(1-\phi\left(\frac{\tau-\mu_0}{\sigma}\right)\right).$$

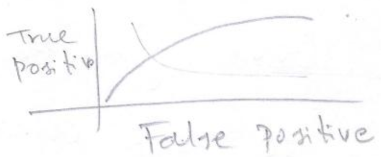
The error is a fn of τ alone.

Now if we vary τ,

$e_0 = P[X \le \tau \mid X \in C1]$ $\quad 1-e_0 = P[X > \tau \mid X \in C1]$

$e_1 = P[X > \tau \mid X \in C_0]$ $\quad 1-e_1 = P[X \le \tau \mid X \in C_0]$

varying τ, makes the point $(e_1, 1-e_0)$ move on a smooth curve in R^2. — This curve is called the ROC.



Metrics of goodness

Accuracy: $\dfrac{\Sigma \text{ True positive} + \Sigma \text{ true negative}}{\Sigma \text{ Total count}}$

Precision: $\dfrac{TP}{TP+FP}$ (proportion of correct detections out of detections) by the classifier.

Recall: $\dfrac{TP}{TP+Fn}$ (proportion of correct detections out of total number of positive cases).

F1-Score: Harmonic mean of P & R.

For a fixed τ, we can estimate e_0 & e_1 from training data.

Hence, varying τ, one can find the 'best' operating point.

Implementing Bayes Classifier: Estimating Class conditional densities.

Problem: Given $\{x_1, x_2, \dots x_m\}$ drawn IID according to some distribution. Estimate the underlying PDF.

Two approaches: i) Parametric: assume the form & estimate the parameter

ii) Non-Parametric: Don't assume the form – modeled as a convex combo of some densities.

Parametric:
- Non-Bayesian (MLE) – Parameter is not a RV; estimate one value.
- Bayesian: Parameter is a RV. Estimate a density.