Evaluating Binary Classifiers

<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>TN</td>
<td>FP</td>
</tr>
<tr>
<td>1</td>
<td>FN</td>
<td>TP</td>
</tr>
</tbody>
</table>

\[ z = \#\text{Neg in GT} \]

TP: True Positive
- GT was "positive" (+1)
- you called it "positive" (+1)

TN: True Negative
- GT was "negative" (class 0)
- you called it "negative"

FP: False Positive
FN: False Negative

\[ \text{TP-Rate} = \frac{\#TP}{\#\text{Pos in GT}} \]
- \( \frac{\#TP}{\#TP + \#FN} \) (aka 'Recall')
- What % of cancer patients did you identify/find?

\[ \text{FP-Rate} = \frac{\#FP}{\#\text{Neg in GT}} \]
- How frequently do you "hallucinate" cancer?
- \( \frac{\#FP}{\#FP + \#TN} \)

Note: Acc = \( \frac{\#TP + \#FP}{\#TP + \#TN + \#FP + \#FN} \)
Let's examine some "dumb" classifiers.

\[ \hat{y}(x) = 1 \quad \forall x \quad "\text{Everybody has cancer!}" \]

**TP-Rate**

\[
\frac{\#TP}{\#Pos-in-Gr} = 1 \quad \text{(100%)}
\]

"Excellent! We found all cancer!"

**FP-Rate**

\[
\frac{\#FP}{\#Neg-in-Gr} = 1 \quad \text{(100%)}
\]

"Oops!"

\[ \hat{y}(x) = 0 \quad \forall x \quad "\text{No-body has cancer}" \]

**FP-Rate**

0

"Yay, no mistakes"

**TP-Rate**

0

"Useless"

In general:

say \( \hat{y}(x) = \text{sign}\left[\frac{\text{confidence}}{\text{Score}(x)} > \text{threshold}\right] \)

E.g. say 10-NN of \( x = 3 \) Pos, 7 Neg

\[ \text{Positive} \quad \text{Score}(x) = \frac{3}{10} = 0.3 \]

Compare this to a threshold \( t \) to decide if your confidence \( > t \)

As threshold \( t \) is varied from 0 \( \rightarrow 1 \), you get a curve:

at \( t = 1 \) \( \hat{y}(x) = 0 \quad \forall x \)

at \( t = 0 \) \( \hat{y}(x) = 1 \quad \forall x \)
ROC Curve

Receiver Operating Characteristics

[Name comes from old-school radio people]

TP-Rate

t=0; TP-Rate=1; FP-Rate=1

(t=0.5, 0.5)

FP-Rate

t=1; TP-Rate=0; FP-Rate=0

Chance/Random Classifier

— Side Note: There is a similar plot called Precision-Recall Curve

where

\[
\text{Precision} = \frac{\# \text{TP}}{\# \text{Pos-Predicted}}
\]

\[
\frac{\# \text{TP}}{\# \text{TP} + \# \text{FP}}
\]

Out of cases we called cancer, how many actually had cancer?
k-NN

\[ \hat{N}_k(x) = \{ \text{indices of } k-\text{NN of } x \text{ in } D \} \]

K-NN predictor

\[ \rightarrow \text{Regression } \hat{y} = g(x) = \frac{1}{k} \sum_{i \in \hat{N}_k(x)} y_i \]

Predict unweighted average of neighbours

\[ \rightarrow \text{Classification } \hat{y} = g(x) = \begin{cases} \arg \max_{c} \sum_{i \in \hat{N}_k(x)} I(y_i = c) \\ \text{class } c \end{cases} \]

= \begin{cases} \arg \max_c \sum_{i \in \hat{N}_k(x)} I(y_i = c) \\ \text{unweighted majority vote} \end{cases} \]
Distances

→ Most common

Euclidean Distance $\| \cdot \|_2$-norm of difference $\mathbf{x}, \mathbf{z} \in \mathbb{R}^d$

$$d(\mathbf{x}, \mathbf{z}) = \left[ \sum_{i=1}^{d} (x_i - z_i)^2 \right]^{1/2}$$

→ hat's generalize this in 2 ways

1. Mahalanobis

New definition

$$d^2(\mathbf{x}, \mathbf{z}) = 10(x_1 - z_1)^2 + (x_2 - z_2)^2$$

↑ deviations in dim1 should be penalized more

In general

$$d^2(\mathbf{x}, \mathbf{z}) = \sum_{i=1}^{d} \sigma_i^2 (x_i - z_i)^2$$

$$= (\mathbf{x} - \mathbf{z})^T \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix} (\mathbf{x} - \mathbf{z})$$

More generally,

$$d^2(\mathbf{x}, \mathbf{z}) = (\mathbf{x} - \mathbf{z})^T A (\mathbf{x} - \mathbf{z})$$

Set $A = \text{Id} \text{dxd} \Rightarrow$ Euc. dist
Note $A \succeq 0$

\[ \text{positive semi-definite} \]

**Definition:** $A = A^T$ symmetric

& $x^T A x \geq 0$ $\forall x \in \mathbb{R}^d$

---

**Other generalization**

Minkowski-distance / $L_p$-norm of difference

\[ d(x, z) = \left[ \sum_{i=1}^d |x_i - z_i|^p \right]^{\frac{1}{p}} \]

\[ p = 2 \equiv \text{Eucl. dist} \]

\[ p = 1 \equiv \text{Manhattan distance} \]

\[ = \sum_{i=1}^d |x_i - z_i| \]

\[ p \to \infty \equiv \text{Max. distance} \]

\[ = \max_i |x_i - z_i| \quad 1 \leq i \leq d \]

**Why? Simple proof**

\[ \lim_{{p \to \infty}} \left[ \sum_{i=1}^d |x_i - d_i|^p \right]^{\frac{1}{p}} \]

Let $j = \text{index of max.-difference}$

\[ = \arg \max_{i=1}^d |x_i - z_i| \]

[For simplicity, assume unique $\arg \max$]
\[
\lim_{p \to \infty} \left[ \sum_{i \neq j} \left( \frac{1}{\|x_i - z_j\|^p} \right)^{1/p} \right]^{1/p}
\]

As \(p \to \infty\),

\[
\left[ \sum_{i \neq j} \left( \frac{1}{\|x_i - z_j\|^p} \right)^{1/p} \right]^{1/p} \to 0
\]

Similarly, \(p = 0\)

\[d(x, z) = \# \text{dums where } x_i \neq z_i\]

**Level Sets**

- \(L_0(z) = \max \{\|x\|_0 \mid 0 = a\}\)
- \(L_1(z) = 1x_1 + 1x_2 = a\)
- \(L_2(z) = \sqrt{x_1^2 + x_2^2} = a\)

For \(p < 1\),

\(L_p(z)\) is a ball.