**ROC Curve**

**Binary classes: P vs. N**

<table>
<thead>
<tr>
<th></th>
<th>P</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>P</td>
<td>Tp</td>
<td>Fp</td>
</tr>
<tr>
<td>N</td>
<td>FN</td>
<td>TN</td>
</tr>
</tbody>
</table>

\[ \text{Tp rate:} \quad \frac{\text{Tp}}{n_p} \quad \text{recall for class P} \]

\[ \text{Fp rate:} \quad \frac{\text{Fp}}{n_N} \quad \text{1-recall for N class} \]

Ideal classifier: \( \text{Fn} = \text{Fp} = 0 \)

\[ \left\{ \begin{align*}
\text{Logistic Regression} & \quad P(Ci|X) \\
\text{SVM} & \quad \text{distance from the hyperplane} \\
\text{Naive Bayes} & \quad P(Ci|X) \\
\end{align*} \right. \]

\( X \): test case

Gradually of predictors,
from the most confidence to less confidence

**ROC curve for a classifier**

Complete picture of the performance
Sort the predictions in decreasing order of signed distance.

<table>
<thead>
<tr>
<th>Point</th>
<th>Signed Distance</th>
<th>Trueclass</th>
<th>PC(P(x))</th>
</tr>
</thead>
<tbody>
<tr>
<td>x_{100}</td>
<td>1000</td>
<td>P</td>
<td>0.7</td>
</tr>
<tr>
<td>x_{50}</td>
<td>500</td>
<td>P</td>
<td>0.8</td>
</tr>
<tr>
<td>x_{5}</td>
<td>10</td>
<td>N</td>
<td>0.71</td>
</tr>
<tr>
<td>x_{200}</td>
<td>1</td>
<td>P</td>
<td>0.7</td>
</tr>
<tr>
<td>x_{201}</td>
<td>1.5</td>
<td>N</td>
<td>0.65</td>
</tr>
</tbody>
</table>

Order of confidence:

P: P
N: N

Threshold for P prediction:

- 0.55
- 0.5
- 0.1

Very strict threshold for P:

(0, 0)

TP: 0
FP: 0
FN: 3
TN: 2

(FPR, TPR) = (0, 0.5)

s = 0.8

All points are in P:

\[ \hat{P} \]

\[ \begin{array}{ccc}
\hat{P} & 3 & 2 \\
0 & 0 & 1 \\
\end{array} \]
Overfitting vs. Generalization:  (memorization vs. learning)

"Simple" vs. "Complex"

Bias of a classifier

Ideal classifier:

Ideal, $A_{UC} = 1$

A: 0.67
B: 0.6

AUC: area under the ROC curve
Linear classifier - high bias

\( f(x) = \hat{y} \) - "shape" of the decision boundary

Bias

\[ \begin{align*}
\text{Linear Regression/SVM} & : \text{high bias} \\
\text{Deep NN} & : \text{low bias}
\end{align*} \]

Robustness to noise or small perturbations

Bias-variance tradeoff
\[ M \leftarrow \text{model} \]
\[ L \leftarrow \text{loss function (SSE)} \]
\[ (Y - \hat{Y})^2 \]

\[ \text{true prediction} \]
\[ \text{expected loss} \]
\[ \text{SS} = \text{loss} \]

\[ \text{true mean} \]
\[ \text{Variance of } Y \]
\[ \text{inherent noise} \]

\[ \mathbb{E} \left[ (Y - \hat{Y})^2 \mid X \right] = \mathbb{E} \left[ (Y - \mu_Y)^2 \mid X \right] + \mathbb{E} \left[ (\hat{Y} - \mu_Y)^2 \mid X \right] \]

\[ \text{Cannot do anything about } \mu_Y \]
\[ \text{This} \]
\[ \text{loss} = \text{avg noise} + \text{avg variance} + \text{avg bias} \]

\[ \text{reduce bias or/and variance} \]

---

**Ensemble classifiers**

**Bagging:** Bootstrap aggregation (effect of reducing variance)

[Diagram of bagging process: Sampling with replacement]
A dataset $D$ is sampled with replacement from $D_{train}$. A model $M_i$ is trained for each sample $D_i$. The predictions $\hat{y}_i$ are then combined using a committee of experts: majority vote for classification and average for regression.
**Boosting:** Adaptive Boosting (AdaBoost)

- **bics reducing effect**

**Bias vs Variance**

**Bagging decision surface**

- "average" effect

**Linear sums**

**D_1**

Initially \( w_i = \frac{1}{n} \)

- Weighted sampling with replacement to obtain \( D_k \)

- \( m_{1} \rightarrow \text{error (e_1)} \)

\[
\begin{align*}
\text{If } e_1 &< 0.5 \quad \text{then} \\
\text{otherwise } &
\end{align*}
\]

- \( \chi_i \leftarrow \text{weight for the classification} = \ln \left( \frac{1 - e_1}{e_1} \right) \)

- \( e_1 = 0.1 \quad \ln(0.9) \)
look at each point which is misclassified

\[ w_i = u_i \cdot \left(1 - \frac{e_i}{e_i}ight) \]

**Example:**

\[ e_1 = 0.5 \]

\[ w_i = w_i \cdot 0.95 \]

Bump up the weight of points misclassified by a "good" classifier.

Renormalize \( w \) to be a probability vector.

Repeat!
weights voting based on weight of the classifier

Decision Trees

greedy recursive partitioning

"axis aligned hyperplanes"

A_1 \leq 10?

Dyes

\begin{array}{c}
4 \times \\
0 \times \\
0 \times \\
0 \times \\
\end{array}

label = X

R_1

Dyes

\begin{array}{c}
x \geq 2 \\
0 \times \\
0 \times \\
\end{array}

label = X

R_2

Dyes

\begin{array}{c}
7 : x \\
0 : 0 \\
\end{array}

Pure partition

label = X

R_3

Dyes

\begin{array}{c}
0 \times \\
0 \times \\
\end{array}

label = 0

R_3

Dyes

\begin{array}{c}
10 \times \\
0 \times \\
\end{array}

not allowed
R₁ : $A₁ \leq 10$ and $A₁ \leq 5$ then $\text{label} = X$

R₂ : $A₁ \leq 10$ and $A₁ > 5$ then $\text{label} = 0$

R₃ : $A₁ > 10$ then $\text{label} = X$

stop R₂
\text{label} = 0