MEASURING CLASSIFIER PERFORMANCE

ERROR COUNTING

Error types in a two-class problem

- **False positives** (type I error): True label is -1, predicted label is +1.
- **False negative** (type II error): True label is +1, predicted label is -1.

We write TP = # true positives, FP = # false positives, TN = # true negatives,
FN = # false negatives

Error rate

$$\text{ER} = \frac{\text{\# wrong predictions}}{\text{\# observations}} = \frac{\text{FP} + \text{FN}}{\text{FP} + \text{FN} + \text{TP} + \text{TN}}$$

Does not distinguish errors between classes.

Relevance

Distinction between error types is crucial e.g. if:

- Classes differ significantly in size
- One type of error has worse consequences than other

MATRIX REPRESENTATION

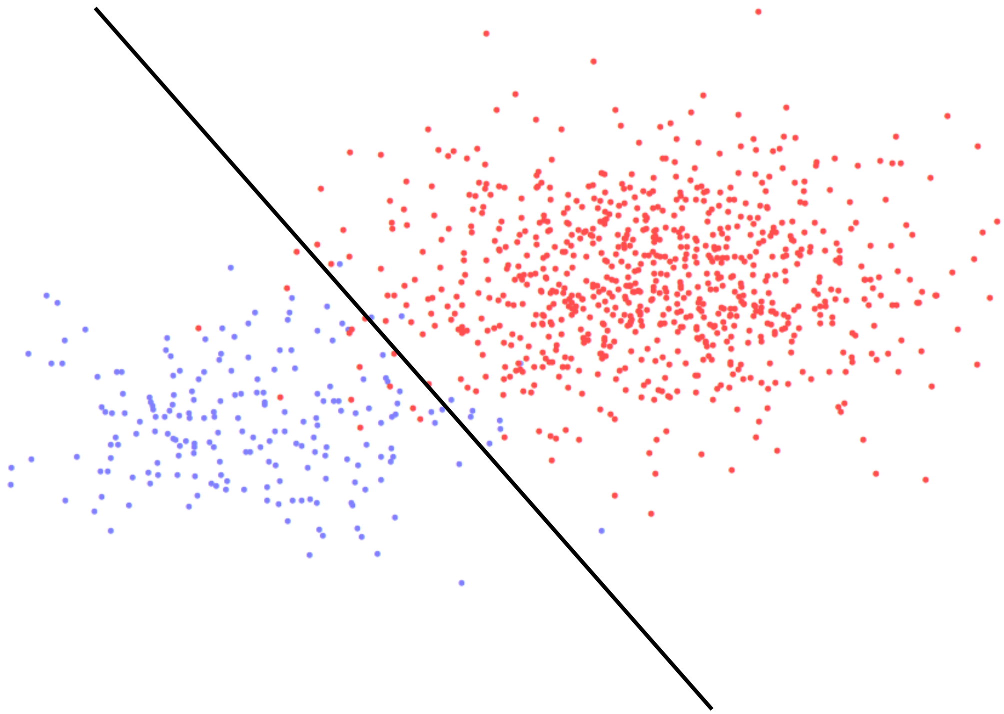
The different types of errors can be summarized in a matrix as

	positive label	negative label
predicted positive	TP/n	FP/n
predicted negative	FN/n	TN/n

where n is the number of observations.

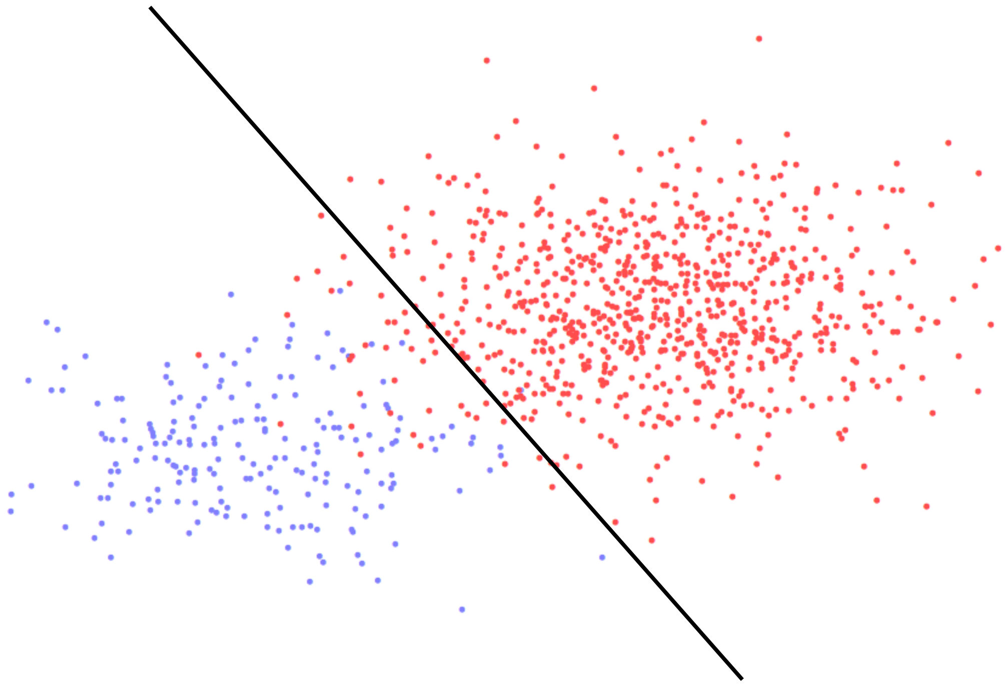
This is called a **confusion matrix** or **contingency table**.

DEPENDENCE ON PARAMETERS



- Suppose a classifier is determined by some parameter θ.
- As we change θ, the number of false positives and false negatives changes.
- We hence have parameter-dependent quantities $\mathrm{TP}(\theta)$, $\mathrm{TN}(\theta)$, etc.

DEPENDENCE ON PARAMETERS



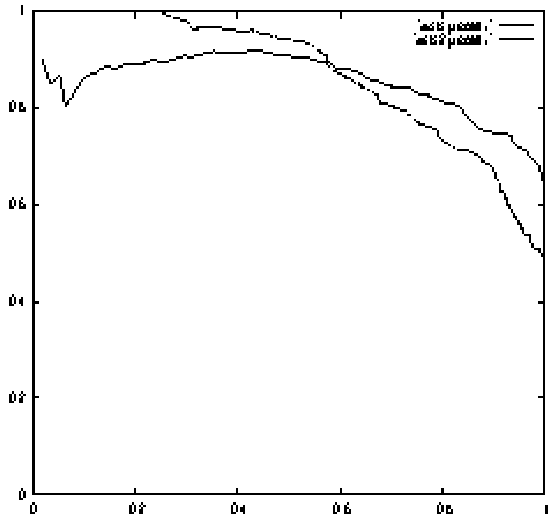
- Suppose a classifier is determined by some parameter θ.
- As we change θ, the number of false positives and false negatives changes.
- We hence have parameter-dependent quantities $\text{TP}(\theta)$, $\text{TN}(\theta)$, etc.

PRECISION AND RECALL

One summary measure of classifier performance are precision and recall:

$$\textbf{Precision}(\theta) := \frac{\text{TP}(\theta)}{\text{TP}(\theta) + \text{FP}(\theta)} \qquad \textbf{Recall}(\theta) := \frac{\text{TP}(\theta)}{\text{TP}(\theta) + \text{FN}(\theta)}$$

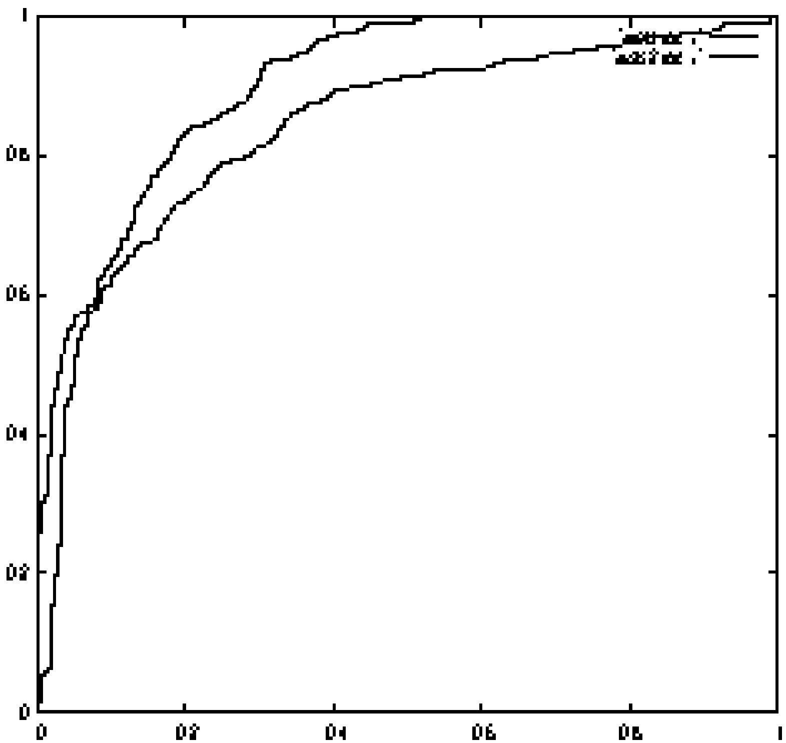
A **precision/recall plot** eveluates precision and recall on validation/test data for a range of different values of θ, and plots precision (vertical axis) against recall (horizontal axis):



- Each point in the plot represents a classifier, for one value of θ.
- Ideally, both precision and recall are high, so "good values" are in the upper right corner.

A plot of the *true positive rate* (TPR) versus the *false positive rate* (FPR) is called a **receiver operating characteristic** (ROC) curve:

$$\text{TPR} = \frac{\text{TP}}{\#\ \text{Positives}} \qquad \text{FPR} = \frac{\text{FP}}{\#\ \text{Negatives}}$$



- "Good" region: Upper left corner. (P/R: Upper *right* corner.)
- Classifier below diagonal (lower left to upper right): Worse than random decision.

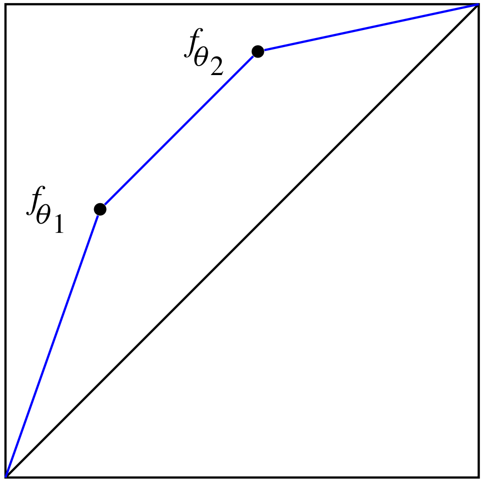
Linear interpolation of classifiers

- Given: Classifiers $f_{\theta_1}, f_{\theta_2}$, interpolation parameter $\lambda \in [0, 1]$.
- Define new classifier f_λ as: Randomly choose output of f_{θ_1} with probability λ, output of f_{θ_2} with probability $1 - \lambda$.

Error rates under interpolation

$$\mathrm{TPR}(f_\lambda) = \lambda \mathrm{TPR}(f_{\theta_1}) + (1 - \lambda)\mathrm{TPR}(f_{\theta_2})$$

The same holds for FPR, ER (but *not* for Precision and Recall).



- ROC plot: Every point represents a classifier performance.
- Consequence: A classifier with performance represented by a point on a straight line between f_{θ_1} and f_{θ_2} in the plot can be constructed by linear interpolation.
- The perfomance values constructable from existing classifiers in this way are called *achievable*.

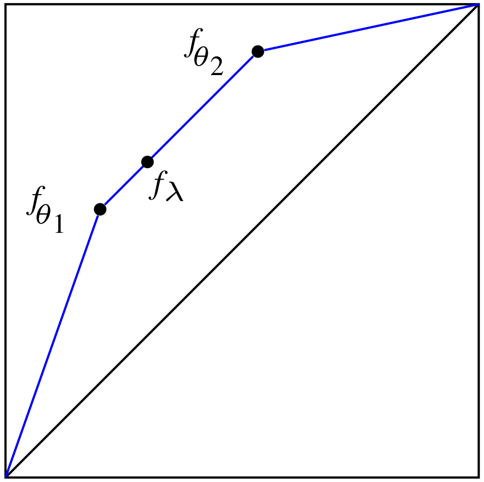
Linear interpolation of classifiers

- Given: Classifiers $f_{\theta_1}, f_{\theta_2}$, interpolation parameter $\lambda \in [0, 1]$.
- Define new classifier f_λ as: Randomly choose output of f_{θ_1} with probability λ, output of f_{θ_2} with probability $1 - \lambda$.

Error rates under interpolation

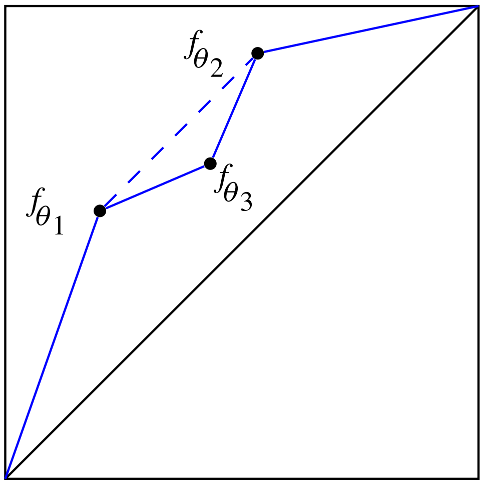
$$\text{TPR}(f_\lambda) = \lambda\text{TPR}(f_{\theta_1}) + (1 - \lambda)\text{TPR}(f_{\theta_2})$$

The same holds for FPR, ER (but *not* for Precision and Recall).



- ROC plot: Every point represents a classifier performance.
- Consequence: A classifier with performance represented by a point on a straight line between f_{θ_1} and f_{θ_2} in the plot can be constructed by linear interpolation.
- The perfomance values constructable from existing classifiers in this way are called *achievable*.

ROC Interpolation: Convex Hull



- Suppose classifiers $f_{\theta_1}, f_{\theta_2}, f_{\theta_3}$ are given:
- If the objective is to optimize ROC performance, f_{θ_3} is worthless.
- We can always obtain a better classifiers by interpolating f_{θ_1} and f_{θ_2}.

In general

- Recall the interpolation formula $\lambda\text{TPR}(f_{\theta_1}) + (1-\lambda)\text{TPR}(f_{\theta_2})$ is a convex combination.
- If $\{f_{\theta_1}, \dots, f_{\theta_k}\}$ are given: Any convex combination of these is achievable.

For given classifiers $\{f_{\theta_1}, \dots, f_{\theta_k}\}$, the convex hull of these classifiers in the ROC plot is achievable.