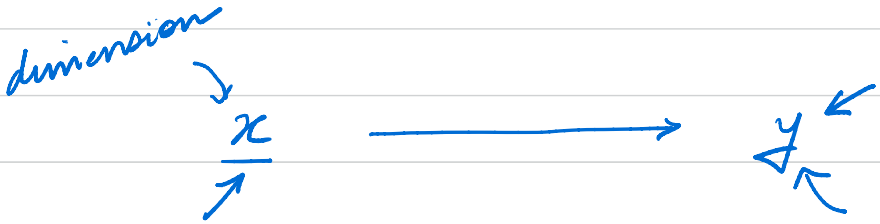
Machine Learning



Goal To learn mapping from $\underline{x}$ to y

Supervised Learning Use training data

τ =

Classification Label y takes on finite # values

If only two values $\{-1, 1\}$ –

If M values $(M > 2)$ –

Classifier uses training data τ to form

f is used to

Measuring Classifier Performance

Loss function $L(\hat{y}, y) = L(f(\underline{x}), y)$

$$\tau = \{(\underline{x}_1, y_1), (\underline{x}_2, y_2), \cdots (\underline{x}_N, y_N)\}$$

Training Error

$$Err_{train}(\tau) =$$

average loss for classifier f over training set τ

For 0-1 loss,

$$Err_{train}(\tau) =$$

It is easy to design f to have $Err_{train}(\tau) = 0$

Example $y \in \{-1, +1\}$, 0-1 loss

$$f(\tilde{\underline{x}}) =$$

Prediction Error and Validation Error

$$Err_{pred} = E\left[L(f(\underline{X}), Y) \mid f \text{ is trained on } \tau\right]$$

Estimate using another labelled data set independent of τ,

$$\hat{Err}_{pred}(\mathcal{V}) =$$

For 0-1 loss,

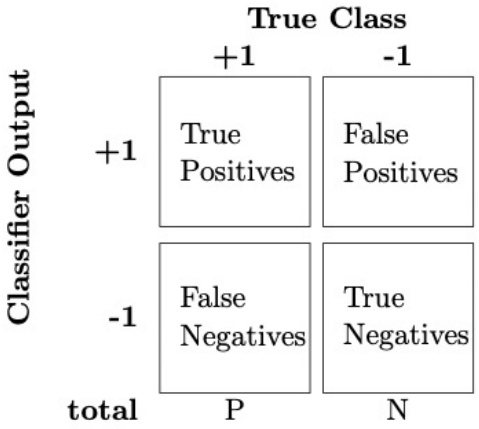
$$Err_{val}(\mathcal{V}) =$$

Displaying Error Performance

- Focus on
- Err_{train} and Err_{val} measure on training and validation datasets.
- More detailed measures of error may be needed for some applications

Confusion Matrix

Focus on special case of classification



P = # data with label

N = # data with label

N_{tot} =

True Positive Rate TPR =

False Positive Rate FPR =

True Negative Rate = specificity =

Precision =

Accuracy =

K-Nearest Neighbor classifier

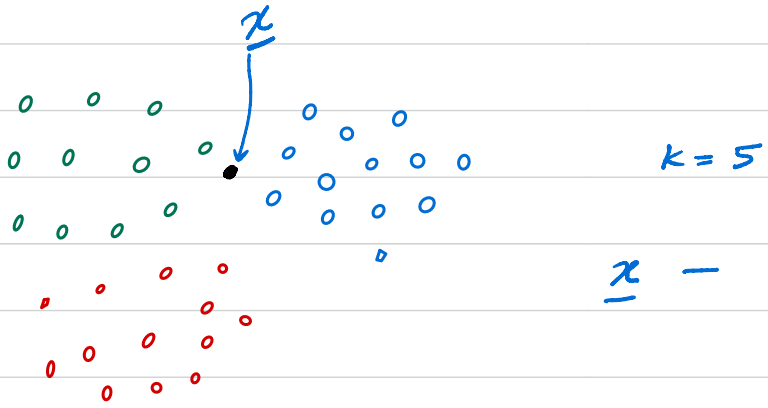
Distance / Dissimilarity Measure

$\Delta(\underline{x}_1, \underline{x}_2)$ measures distance between feature vectors $\underline{x}_1$ and $\underline{x}_2$

Examples: 1) $\Delta(\underline{x}_1, \underline{x}_2) =$

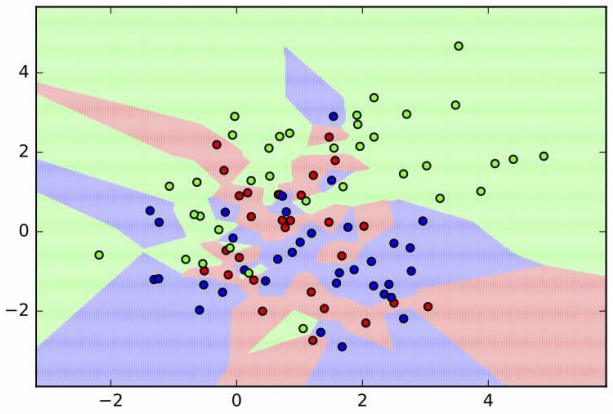
2) $\Delta(x_1, x_2) =$

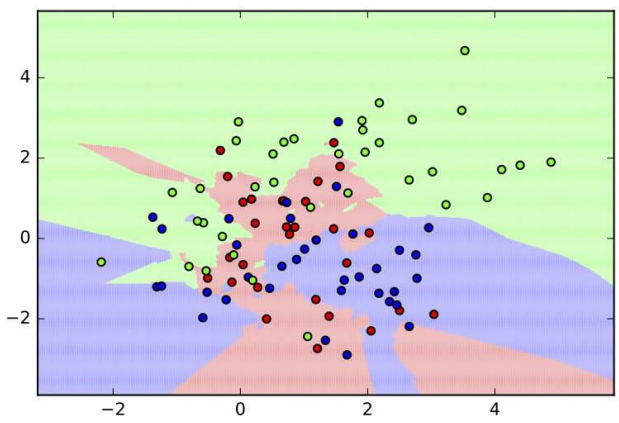
K-NN



For each $\underline{x}$ pick k nearest neighbors using $\Delta(\cdot,\cdot)$ and choose label as most label.

For $k=1$, algorithm is simply called NN classifier





Binary Bayes Classifier $\underline{x} \longrightarrow y \in \{+1, -1\}$

Suppose we know:

$$\pi_y \triangleq p(y) = P\{Y = y\}, \quad y = +1, -1$$

and $p(\underline{x} \mid y), \quad y = +1, -1.$

Then, binary classification problem is equivalent to binary hypothesis testing:

Joint distribution

For 0-1 loss,

$Err_{pred} =$

MAP Rule

$$\frac{p(\underline{x} \mid 1)}{p(\underline{x} \mid -1)} \underset{\text{say } -1}{\overset{\text{say } 1}{\gtrless}} \frac{\pi_{-1}}{\pi_1} \equiv$$

$\hat{y}_{MAP} =$

MAP rule is also called

$\hat{y}_{Bayes} = f_{Bayes}(\underline{x})$

f_{Bayes}