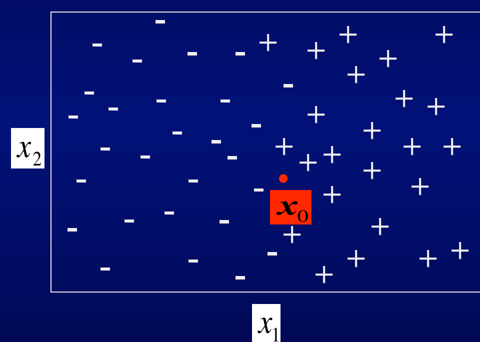
# Classification Problem

- Given $\{(\boldsymbol{x}_n, y_n)\}_{n=1}^{N}$ $\boldsymbol{x}_n \in \Re^q$ $y_n \in \{+, -\}$




- Predict class label of a given query $\boldsymbol{x}_0$

# Classification Problem

- Unknown probability distribution $P(\boldsymbol{x}, y)$

- We need to estimate:

$$P(+ \mid \boldsymbol{x}_0) \equiv f_+(\boldsymbol{x}_0)$$

$$P(- \mid \boldsymbol{x}_0) \equiv f_-(\boldsymbol{x}_0)$$

# The Bayesian Classifier

- Loss function: $\lambda(j \mid k)$
- Expected loss (conditional risk) associated with class *j*:

$$R(j \mid \boldsymbol{x}) = \sum_{k=1}^{J} \lambda(j \mid k)\, P(k \mid \boldsymbol{x})$$

- Bayes rule:

$$j^* = \arg\min_{1 \le j \le J} R(j \mid \boldsymbol{x})$$

- Zero-one loss function:

$$\lambda(j \mid k) = \begin{cases} 0 & \text{if } j = k \\ 1 & \text{if } j \ne k \end{cases}$$

$$j^* = \arg\max_{1 \le j \le J} P(j \mid \boldsymbol{x})$$

**Bayes rule**

# The Bayesian Classifier

$$j^* = \arg\max_{1 \le j \le J} P(j \mid \boldsymbol{x})$$

- Bayes rule achieves the minimum error rate
- How to estimate the posterior probabilities:

$$\left\{P(j \mid \boldsymbol{x})\right\}_{j=1}^{J}$$

$$\hat{j}(\boldsymbol{x}) = \arg\max_{1 \le j \le J} \hat{P}(j \mid \boldsymbol{x})$$

# Density estimation

- Use Bayes theorem to estimate the posterior probability values:

$$P(j \mid \boldsymbol{x}) = \frac{p(\boldsymbol{x} \mid j)P(j)}{\sum_{k=1}^{J} p(\boldsymbol{x} \mid k)P(k)}$$

$p(\boldsymbol{x} \mid j)$ is the probability density function of $\boldsymbol{x}$ given class $j$

$P(j)$ is the prior probability of class $j$

# Naïve Bayes Classifier

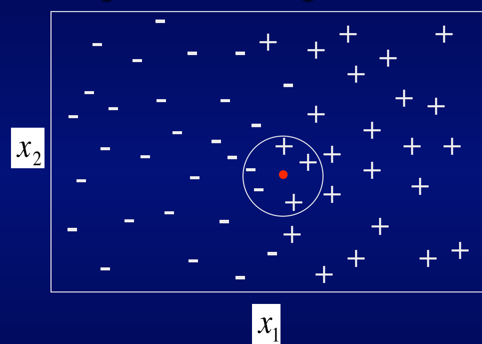
- Makes the assumption of independence of features given the class:

$$p(\boldsymbol{x} \mid j) = p(x_1, x_2, \cdots, x_q \mid j) = \prod_{i=1}^{q} p(x_i \mid j)$$

- The task of estimating a $q$-dimensional density function is reduced to the estimation of q one-dimensional density functions. Thus, the complexity of the task is drastically reduced.
- The use of Bayes theorem becomes much simpler.
- Proven to be effective in practice.

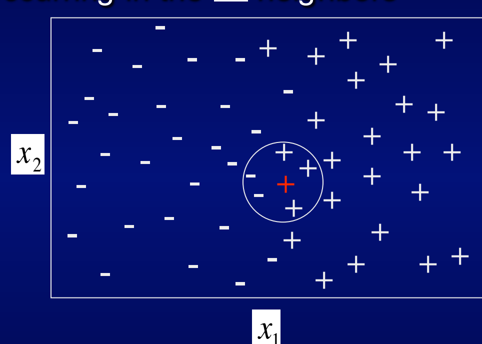
# Nearest-Neighbor Methods

- Predict the class label of $x_0$ as the most frequent one occurring in the $K$ neighbors



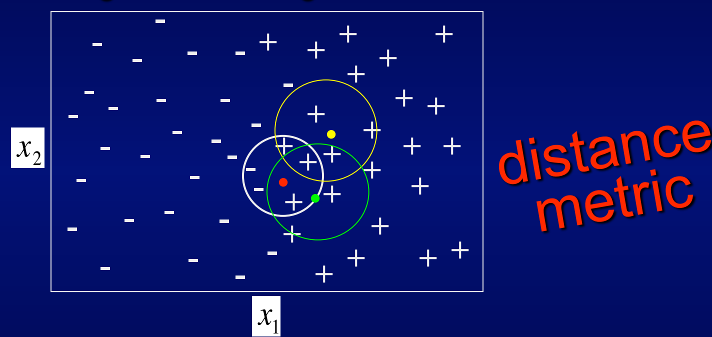

# Nearest-Neighbor Methods

- Predict the class label of $x_0$ as the most frequent one occurring in the $K$ neighbors



# Nearest-Neighbor Methods

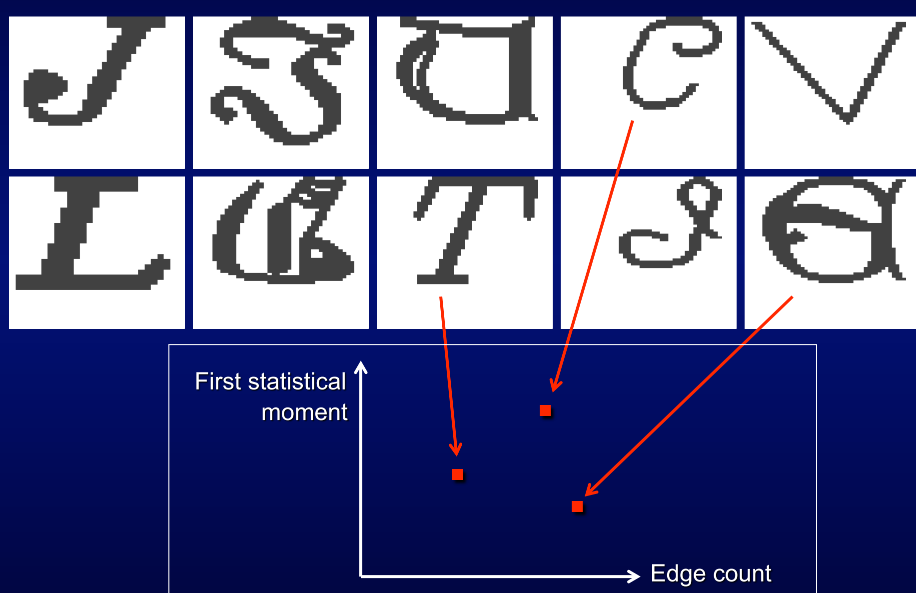
- Predict the class label of $x_0$ as the most frequent one occurring in the $K$ neighbors




Basic assumption:

$$f_+(\boldsymbol{x}+\delta\boldsymbol{x}) \approx f_+(\boldsymbol{x})$$
$$f_-(\boldsymbol{x}+\delta\boldsymbol{x}) \approx f_-(\boldsymbol{x})$$
for small $\|\delta\boldsymbol{x}\|$

# Example: Letter Recognition



# Asymptotic Properties of K-NN Methods

$$\lim_{N\to\infty} \hat{f}_j(\boldsymbol{x}) = f_j(\boldsymbol{x})$$

if $\lim_{N\to\infty} K = \infty$ and $\lim_{N\to\infty} K/N = 0$

- The first condition reduces the variance by making the estimation independent of the accidental characteristics of the *K* nearest neighbors.
- The second condition reduces the bias by assuring that the *K* nearest neighbors are arbitrarily close to the query point.

# Asymptotic Properties of K-NN Methods

$$\lim_{N\to\infty} E_1 \leq 2E_\infty$$

$E_1 \equiv$ classification error rate of the 1-NN rule

$E_\infty \equiv$ classification error rate of the Bayes rule

In the asymptotic limit no decision rule is more than twice as accurate as the 1-NN rule

## Finite-sample settings

- How well the 1-NN rule works in finite-sample settings?

- If the number of training data *N* is large and the number of input features *q* is small, then the asymptotic results may still be valid.

- However, for a moderate to large number of input variables, the sample required for their validity is beyond feasibility.

## Curse-of-Dimensionality

- This phenomenon is known as the ***curse-of-dimensionality***
- It refers to the fact that in high dimensional spaces data become extremely sparse and are far apart from each other
- **It affects *any* estimation problem with high dimensionality**