Classification

\[D = \{(x^{(1)}, y^{(1)}), (x^{(2)}, y^{(2)}), \ldots, (x^{(N)}, y^{(N)})\}\]

\[\forall i \quad x^{(i)} \in \mathbb{R}^M \quad \text{← features (input)}\]
\[\forall i \quad y^{(i)} \in \{1, 2, \ldots, L\} \quad \text{← labels (output)}\]

Def: Binary Classification

Above where \(y^{(i)} \in \{0, 1\}\)
\(y^{(i)} \in \{+1, -1\}\)

\[\hat{y} = [y^{(1)}, \ldots, y^{(N)}] \quad \hat{x} = [x^{(1)\text{T}}, \ldots, x^{(N)\text{T}}] \quad x^{(i)} = [x_1^{(i)}, \ldots, x_M^{(i)}]\]

Ex: 2D Bin. Class

Linear Decision Boundary

**Nonlinear Decision Rule**

**Def:** Decision Rule for Bin. Class.

**Function** $h : \mathbb{R}^M \rightarrow \{0,1\}$

Given $\hat{x}(\text{test})$, we predict $\hat{y} = h(\hat{x}(\text{test}))$

---

**k-Nearest Neighbors**

**Def:** (English) "Assign most common label among $k$ neighbors in $D$ that are nearest to $\hat{x}(\text{test})$".

**Def:** (Picture) $k = 3$
**Def:** Algorithm

**Inputs**
- \( k = \# \) nearest neighbors
- \( d = \) distance function
- \( D = \) data

**1. Train:**
- Given train ex. s \( D \)
- Store \( D \)
- Return

**2. Predict:**
- Given test ex. \( \hat{x} \) (test)
- Find \( k \) nearest neighbors

\[ I = \{ n_1, n_2, \ldots, n_k \} \]

\( \forall j, n_j \in I, \ n_j \in \{ 1, 2, \ldots, N \} \)

\[ s.t. \quad d(\hat{x} \text{ (test)}, \hat{x}(n_j)) \leq d(\hat{x} \text{ (test)}, \hat{x}(i)) \]

\( n_j \in I, \ i \not\in I \)

- Return majority label

\[ \hat{y} = h(\hat{x} \text{ (test)}) = \arg\max_{l \in \{0,1\}} \sum_{n_j \in I} \mathbb{1}(l, y(n_j)) \]

Indicator fn.

\( \mathbb{1}(u,v) = 1 \) if \( u = v \)

\( = 0 \) otherwise

**Remarks:** Distance fn. \( d \) is usually Euclidean

\[ d(\hat{u}, \hat{v}) = \sqrt{\sum_{M=1}^{M} (u_m - v_m)^2} \]

Other fns. are fine

\[ d(\hat{u}, \hat{v}) = \sum_{M=1}^{M} |u_m - v_m| \]

\[ y(\hat{x}(t)) = h(\hat{x}(t)) = 0 \]

\( I = \{ 4, 7, 10, 3 \} \)