Non-parametric: density estimation (histogram)

Numeric attribute \( A_j \)
for a single attribute

\[ f(x_j) \]

Normal vs histogram

\[ f(x_j) \]

\[ \frac{3}{6} \]
\[ \frac{1}{6} \]
\[ \frac{2}{6} \]

\[ \frac{1}{3} \]
\[ \frac{1}{3} \]

\( d \) dims

2 bins per dim

\( 2^d \) cells
$2^{1000}$ cells, $j = 1000$

$\approx 2^{60}$ atoms in universe

$k$-Nearest Neighbors (density estimation)

No training

$\hat{x} \in$ test point

What is its class?

$1$NN?

$[\text{1st nearest neighbor}]$

$\text{NN}(\hat{x}) = \arg \min_{x_j} \| \hat{x} - x_j \|$

Predicts class in the

$\hat{y} = \text{class}(x_j)$

$\hat{y} = 0$

$1$NN

$k$-NN:

Select test point

$N_k(\hat{x})$: set of $k$ points nearest to $\hat{x}$ in $D$
\[ N_k(\mathbf{x}) : \text{set of } k \text{ points nearest to } \mathbf{x} \in D \]

\[ \hat{y} = \text{majority class in } N_k(\mathbf{x}) \]

\[ \hat{P}(c_i | \mathbf{x}) = \frac{\text{how many training pts are in } c_i}{N_k(\mathbf{x})} \]

---

**B-trees, R-trees, Oct-trees**

*Spaced indexing methods*

---

1NN

*Coarse approx*

\( \hat{K}-\text{NN} \)

What's the right \( k \)?

---

Support Vector Machines (SVM)

\[ \uparrow \]

Structural Risk Minimization

Large margin classifier

\[ \downarrow \]

Hyperplanes
SVM will solve for $h(x)$ hyperplane.

Binary classification

Large margin classifier

Support vectors.
A hyperplane $h(x)$ is defined as $h(x) = \omega^T x + b = 0$, where $\omega$ and $b$ are parameters. The region where $h(x) < 0$ is called the half-space, and the region where $h(x) > 0$ is the other half-space.

The margin of a point $\mathbf{x}$ is the signed distance of the point to the hyperplane, defined as $\gamma = \frac{h(x)}{||\omega||} = \frac{\omega^T x + b}{||\omega||}$. The sign function $\text{sign}(h(x))$ determines whether a point is on the positive or negative side of the hyperplane:

$$\text{sign}(h(x)) = \begin{cases} +1 & \text{if } h(x) > 0 \\ -1 & \text{if } h(x) < 0 \end{cases}$$
\[ h(x) > 0 \]

\[ y = \text{sign}(h(x)) \]

- \[ h(x) > 0 \quad y = +1 \]
- \[ h(x) < 0 \quad y = -1 \]
$$\delta_x = \text{normal distance} = \frac{y \cdot h(x)}{||w||}$$ (unsigned)

Margin $\delta h(x) = \text{minimum distance for any point in D to } h(x)$

Margin $\delta^+ = \min_{x_j} \left\{ \frac{y_j \cdot h(x_j)}{||w||} \right\}$

Objective is maximize $\delta^+$ (the margin)

Find $w \in \mathbb{R}^n$, such that

$$\max \left\{ \delta^+ \right\} = \max \left\{ \min_{x_j} \left\{ \frac{y_j \cdot h(x_j)}{||w||} \right\} \right\}$$

$$\max \left\{ \min_{x_j} \left\{ \frac{y_j \cdot (w \cdot x_j + b)}{||w||} \right\} \right\}$$

Can be made arbitrarily small by choosing large $w$

$$\frac{v_3}{2} = \begin{pmatrix} 1 & 1 & 1 \\ 2 & 2 & 2 \end{pmatrix}^T \begin{pmatrix} 1 \\ 2 \\ 2 \end{pmatrix} = \begin{pmatrix} 100 \\ 200 \end{pmatrix}$$
\[ s_x = \frac{y_j \left( \hat{\omega}^T \hat{x}_j + b \right)}{\| \hat{w} \|} = \frac{1}{\| \hat{w} \|} \]

**Canonical hyperplane:**

- Closer point is "set" at distance 1
- Unit of measurement

Margin of any hyperplane is \( \frac{1}{\| \hat{w} \|} \)

Find \( \hat{w}, \hat{b} \)

\[
\max \left\{ \frac{1}{\| \hat{w} \|} \right\}
\]

Subject to the constraint that the closer point is at

\[
\text{distance } \frac{1}{\| \hat{w} \|}
\]

Closest point satisfies the equation

\[ y_j \left( \hat{\omega}^T \hat{x}_j + b \right) = 1 \]

Really in different constraint, 1 per point

\[ y_i \left( \hat{\omega}^T \hat{x}_i + b \right) \geq 1 \quad \forall x_i \in D \]
\[
\begin{align*}
&\text{max} \left\{ \frac{1}{\|w\|} \right\} \\
\text{s.t.} \quad &y_i (\mathbf{w}^T \mathbf{x}_i + b) \geq 1 \quad \forall x_i \in \mathcal{D}
\end{align*}
\]

\[
\frac{1}{\|w\|} = \sqrt{w_1^2 + w_2^2 + \cdots + w_d^2}
\]

\[
\begin{align*}
&\text{min} \left\{ \|w\| \right\} \\
&\text{s.t.} \quad y_i (\mathbf{w}^T \mathbf{x}_i + b) \geq 1
\end{align*}
\]

\[
\begin{align*}
&\text{min} \left\{ \frac{1}{2} \|w\|^2 \right\} \\
&\text{s.t.} \quad y_i (\mathbf{w}^T \mathbf{x}_i + b) \geq 1 \quad \forall x_i
\end{align*}
\]

\[
\text{Convex Qudratic Program} \quad \Rightarrow \quad \text{Unique Solution, globally optimal solution}
\]

\[
\text{Obj: } \min \left( \frac{1}{2} \|w\|^2 \right)
\]
\[ y_i \left( \omega^T x_i + b \right) \geq 1 \quad \forall x_i, \ i = 1, \ldots, n \]

Lagrange multipliers \( \alpha_i \geq 0 \), one per constraint

\[
J = \min \left\{ \frac{1}{2} \|w\|^2 - \sum_{i=1}^{n} \alpha_i \left( y_i (\omega^T x_i + b) - 1 \right) \right\}
\]

final "primal" SVM objective

\[
\frac{\partial J}{\partial w} = \omega - \sum_{i=1}^{n} \alpha_i y_i x_i = 0
\]

\[
\Rightarrow \quad \lambda = \sum_{i=1}^{n} \alpha_i y_i x_i = 0
\]

\[
\alpha_i \geq 0
\]

\[\hat{\omega} = \text{linear combination of } x_i \text{'s (y_i's)} \]

\[
\frac{\partial J}{\partial b} = \sum \alpha_i y_i = 0
\]

Optimality condition:

\[\alpha_i \left( y_i \left( \omega^T x_i + b \right) - 1 \right) = 0 \quad \text{for the closest points} \]

also called support vectors