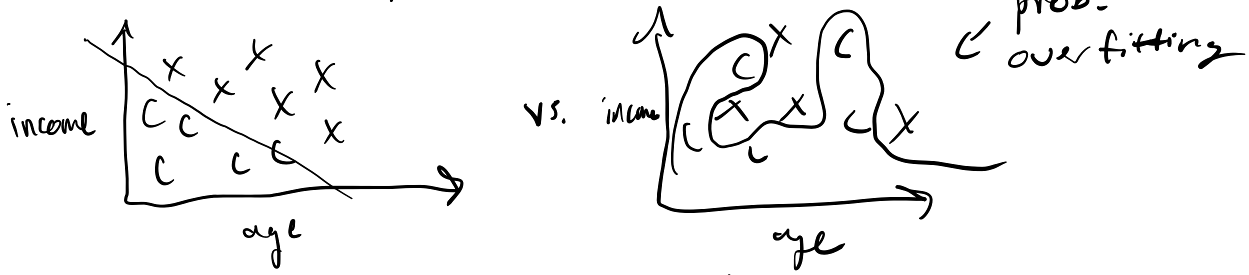
Linear Classifiers. Decision functions / boundaries
The centroid method. Perceptrons.

1.) Classifiers

- given a sample of n observations, d features (predictors)
  - some are class C: some are not
    - ↳ class "defaulted" for bank loans
    - ↳ predict default based on income / age

point in d-dim space (d×n matrix)



overfitting: when decision boundaries fit sample points so well it doesn't fit test data well.

decision function $f(x)$ over feature space (d-dim)

$\mathbb{R}^d \to \mathbb{R}$ $\quad f(x) > 0$ if $x \in C$

$f(x) \leq 0$ if $x \notin C$

decision boundary $\{x \in \mathbb{R}^d : f(x) = 0\}$ isosurface w/ isovalue 0

linear classifier: decision boundary is a line / plane

$$* \cos\theta = \frac{X}{\|X\|_2} \cdot \frac{Y}{\|Y\|_2}$$

$f(x) = w^T x + \alpha$ $\qquad H = \{x : w^T x = -\alpha\}$

$w \perp H$

$w^T(y - x) = 0$

$-\alpha - (-\alpha) = 0$

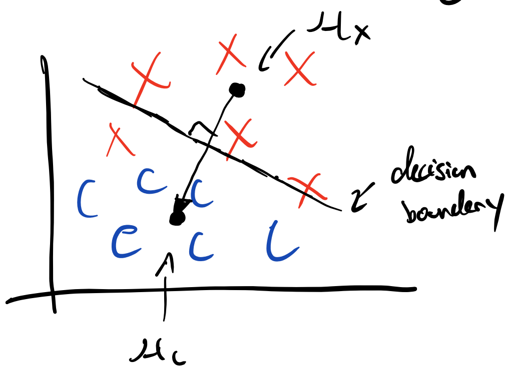
If w is unit vector, $w^T x + \alpha$ is signed distance
- (+) if w side of H, (−) else
- dist. H to origin $= \alpha$
- $\alpha = 0 \iff H$ passes through origin

coeffs. (w, α) are <u>weights</u> of lin. classifier.

Centroid method:

compute μ of class C (μ_c) and class not C (μ_x)

$$f(x) = (\mu_c - \mu_x)^T x - (\mu_c - \mu_c)^T \left(\frac{\mu_c + \mu_x}{2}\right)$$



Perceptron algorithm (Rosenblatt, 1957)

- slow, correct for lin. sep points
- gradient descent

- n pts. $X_1 \; X_2 \ldots X_n$ $\quad y_i = \begin{cases} 1 & \text{if } x_i \in C \\ -1 & \text{else} \end{cases}$

for now, $\alpha = 0$ (boundary through origin)

$x_i^T w \geq 0$ if $y_i = 1$

$x_i^T w < 0$ if $y_i = -1$

$y_i X_i^T w \geq 0$

- risk function R is positive if no contraints violated
- optimization to minimize R

loss function $$L(z, y_i) = \begin{cases} 0 & \text{if } y_i z \geq 0 \\ -y_i z & \text{otherwise} \end{cases}$$

$$R(w) = \frac{1}{n} \sum_{i=1}^{n} L(X_i^T w, y_i) = \frac{1}{n} \sum_{i \in V} -y_i X_i^T w$$

V is misclassified points

Solve: $$\min_{w} R(w)$$