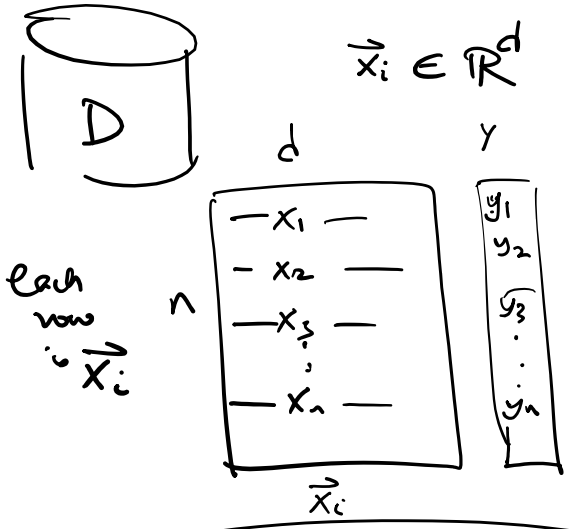
# Pretrained models for sequences

↳ logistic regression

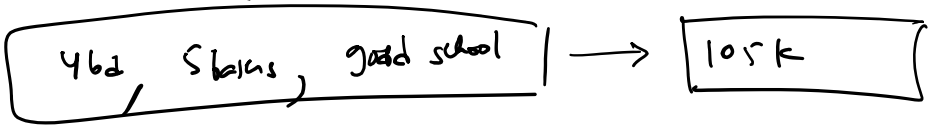
↳ linear regression

---




d-dimensional vector

| | $A_1$ | $A_2$ | $A_d$ |
|---|---|---|---|
| $x_1$ = | 1 | 2 | 1.5 |
| $x_2$ | 0 | 1 | 2 |
| $x_3$ | 0.5 | 0.2 | 0.9 |
| | | | |




$A_1$ $A_2$ . . $A_d$ ← attributes / variables

independent attributes

$y$ ← response, dependent attribute

$$\hat{y} = f(A_1, A_2, \ldots, A_d)$$

$\hat{y}$: predicted response

f : very simple (linear regression)

f : very complex (deep CNN)

$$\hat{y} = w_0 + w_1 \cdot A_1 + w_2 A_2 + \cdots + w_d A_d$$ ← linear model

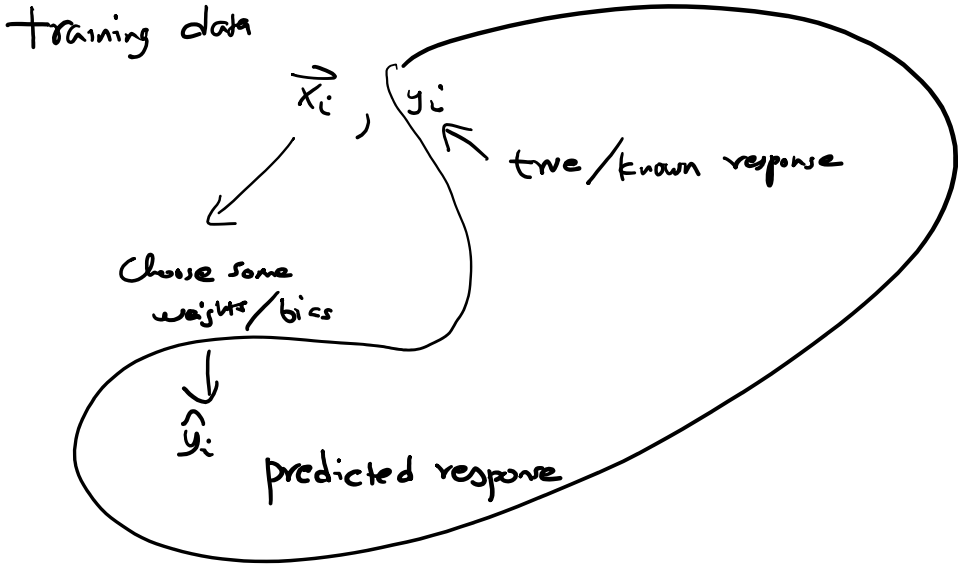
$w_0$: bias

$w_1, w_2, \ldots, w_d$ are unknown

regression coefficients

weights

$w_0, w_1, \ldots, w_d$ ← parameters




loss/error function: squared error/loss

$$\text{objective} = (y_i - \hat{y_i})^2$$

$$SSE = \sum_{i=1}^{n} (y_i - \hat{y_i})^2$$

↑ sum of squared errors

Task: find the parameters $w_0, w_1, w_2, \ldots, w_d$ such that we minimize SSE

Task: find ... such that we minimize SSE

$$\min_{w_0, w_1, \ldots, w_d} J = \sum_{i=1}^{n} \left( y_i - w_0 - w_1 x_{i1} - w_2 x_{i2} - \cdots - w_d x_{id} \right)^2$$

$d+1$ real values

$$\vec{x}_i = \left( x_{i1}, x_{i2}, \ldots, x_{id} \right)$$

Optimization method: SGD
Stochastic gradient Descent

$$\nabla_{w_i} = \frac{\partial J}{\partial w_i}$$

Start from any random value

$w_0 = 1$

$$w_0 = w_0 - \eta \cdot \nabla_{w_0}$$ ⟵ SGD update rule

$\eta$: stepsize / learning rate