- centered coordinates
- gradient descent

Linear Regression

X data matrix $N \times (K+1)$
(last col is all 1's)

Y target data $N \times 1$

M $(K+1) \times 1$

Find M:

$$X^T(\underset{N\times 1}{Y} - \underset{N \times (K+1)\,(K+1)\times 1}{XM}) = 0$$

$(K+1) \times N$

$$(\underbrace{X^T X}_{D} M = X^T Y)$$

$$M = D^{-1}(X^T Y)$$

$$Y - XM = \begin{bmatrix} y_i - \sum_{i=1}^{K+1} x_i m_i \\ \vdots \\ \vdots \end{bmatrix}$$

$$X^T = \begin{pmatrix} \\ \\ 1\ 1\ 1\ 1 \cdots 1 \end{pmatrix}$$

$$X^T(Y - XM) = \begin{bmatrix} \vdots \\ \sum_{i=1}^{N} \left[y_i - \sum_{j=1}^{K+1} x_{ij} m_j \right] \end{bmatrix}$$

y_n $= \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}$

$$\sum \begin{matrix} y_1 - x_{11}m_1 - x_{12}m_2 \cdots - x_{1k}m_k - b \\ y_2 - x_{21}m_1 - \cdots - x_{2k}m_k - b \\ \vdots \end{matrix}$$

$$\sum y_i - \left(\sum x_{i1}\right)m_1 + \sum(x_{i2})m_2 + \ldots N b$$ Eq

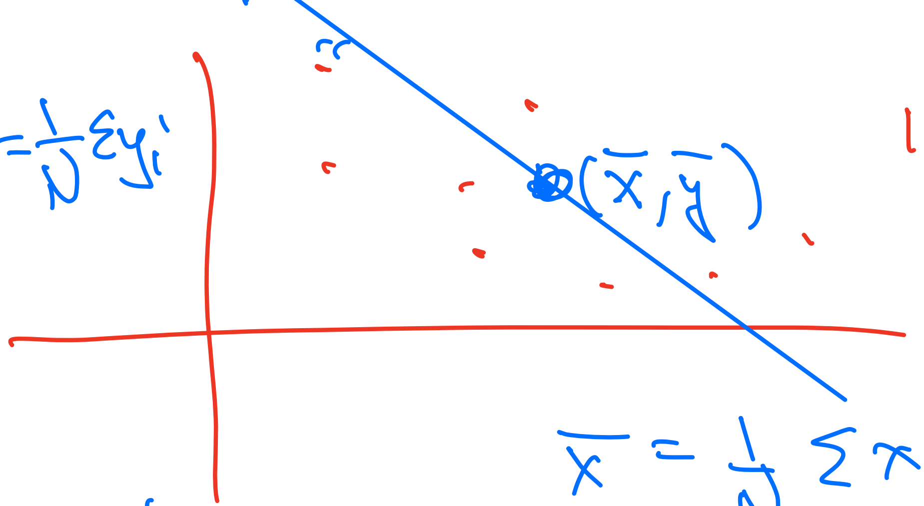
$$\frac{1}{N}\sum y_i = \left(\frac{1}{N}\sum x_{i1}\right)m_1 + \frac{1}{N}\left(\sum x_{i2}\right)m_2 + \ldots + b = 0$$

$$b = \bar{y} - \sum_{i=1}^{k} \bar{x}_i m_i$$

$$\begin{array}{ccccc} & y & x_1 & x_2 & \\ & \text{mpg} & \text{disp} & \text{hp} & \ldots \\ 384 \text{ rows} & \big(\, | & | & | & \big) \end{array}$$

$\bar{y} = \overline{\text{mpg}}$

$\bar{y} = \frac{1}{N}\sum y_i$



(x_i, y_i)

$\bar{y}$

$b = y - mx$

$\bar{y} - m\bar{x}$

$\bar{x} = \frac{1}{N}\sum x_i$

$y = mx + b$

<u>First Step</u>:

Shift coordinates so means of target and the features are all <u>zero</u>.

<u>Centered</u> <u>coordinates</u>.

Know: $b = 0$

No need to add a column of ones

Machine Learning (supervised)

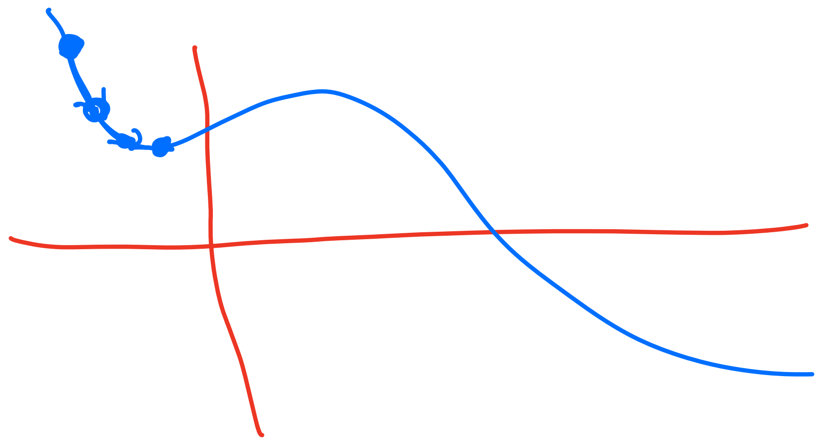
Data ✓ targets

model (unknown parameters)

Goodness of fit (loss function)

opt Find parameters where loss function is minimized

Gradient Descent



$y = f(x)$

Numerical iterative methods

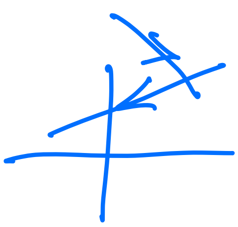
$y = f(x)$ Find local minimum.

$x_0 \longleftrightarrow$ initial point

calculate $f'(x_0) \longleftrightarrow$ "slope"

$+$ ~~go left~~ go left (subtract from x_0)

$-$ right (add to x_0)



$x_1 = x_0 - \lambda \underline{f'(x_0)}$

$\lambda \longleftrightarrow$ "learning rate"

"$f'(x_0)$" $\approx \dfrac{f(x_0+\varepsilon)-f(x_0)}{\varepsilon}$

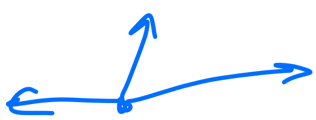
iterate $x_2 = x_1 - \lambda f'(x_1)$

$\vdots$

$$z = f(x_1, x_2, \ldots, x_n)$$

$(x_1, \ldots, x_n) = (a_1, \quad , a_n)$ critical point

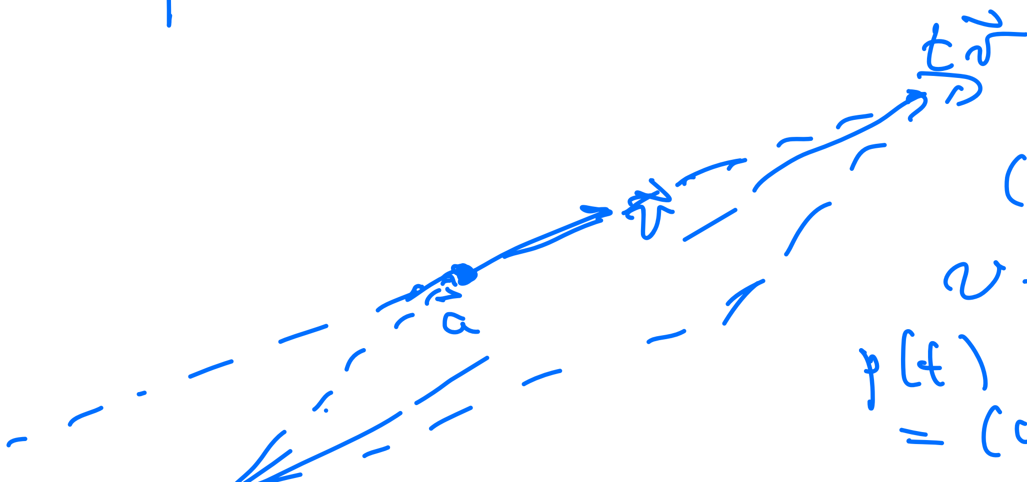
where to go?



$v = (v_1, \ldots, v_n)$

$(a_1 \ldots a_n)$

$$p(t) = (a_1, \ldots, a_n) + t(v_1, \ldots, v_n)$$



(a_1, a_2)

$v = (v_1, v_2)$

$p(t)$
$= (a_1 + t v_1, a_2 + t v_2)$

~~f~~ $f(x_1, \ldots, x_n)$

Look at

$f(a_1 + tv_1, a_2 + tv_2, \ldots, a_n + tv_n)$

$f(\vec{a} + t\vec{v})$

$\frac{df}{dt}$ rate of change in height w.r.t time as you travel from $\vec{a}$ with velocity $\vec{v}$ in a line.

$$\frac{d}{dt} f(a_1 + tv_1, a_2 + tv_2, \ldots, a_n + tv_n)$$

$=$

$$\frac{df}{dt} = \frac{\partial f}{\partial x_1}\frac{\partial x_1}{\partial t} + \frac{\partial f}{\partial x_2}\underbrace{\frac{\partial x_2}{\partial t}}_{v_2} + \ldots + \frac{\partial f}{\partial x_n}\underbrace{\frac{\partial x_n}{\partial t}}_{v_n}$$

$x_1 = a_1 + tv_1$

$$\frac{\partial x_1}{\partial t} = v_1$$

$$= \left[\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \ldots, \frac{\partial f}{\partial x_n}\right] \cdot [v_1, \ldots, v_n]$$

$$\frac{df}{dt} = (\nabla f) \cdot \vec{v}$$

Directional Derivative of f in direction of v

$$(D_v f)(a) = [(\nabla f)(a)] \cdot \vec{v}$$

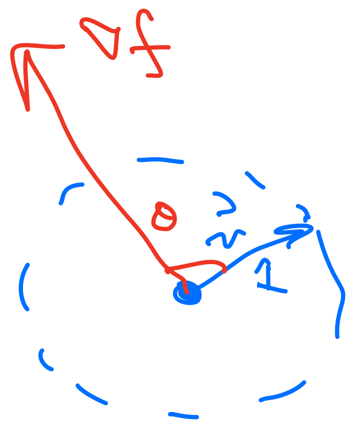
$$(D_v f)(a) = |\nabla f \cdot \vec{v}|$$

$$= \|\nabla f\| \cdot \|\vec{v}\| \cos\theta$$

$$\|\vec{v}\| = 1$$

f increases most rapidly in direction of ∇f

decreases most rapidly in direction of $-\nabla f$



Given f

$x^{(0)}$ = initial guess $(a_1, a_2, \ldots, a_n)$

Compute ∇f at x_0 (vector)

$x^{(1)} = x^{(0)} - \nu(\nabla f)$

repeat this

Linear Regression

minimize: $\sum \left(y_i - \sum x_{ij} m_j\right)^2 = E$

$\nabla_m E = 2X^T(Y - Xm)$ m variable

$m^{(i+1)} = m^{(i)} - \lambda\left[2X^T(Y - Xm^{(i)})\right]$

$m^{(0)} \leftarrow$ random guess