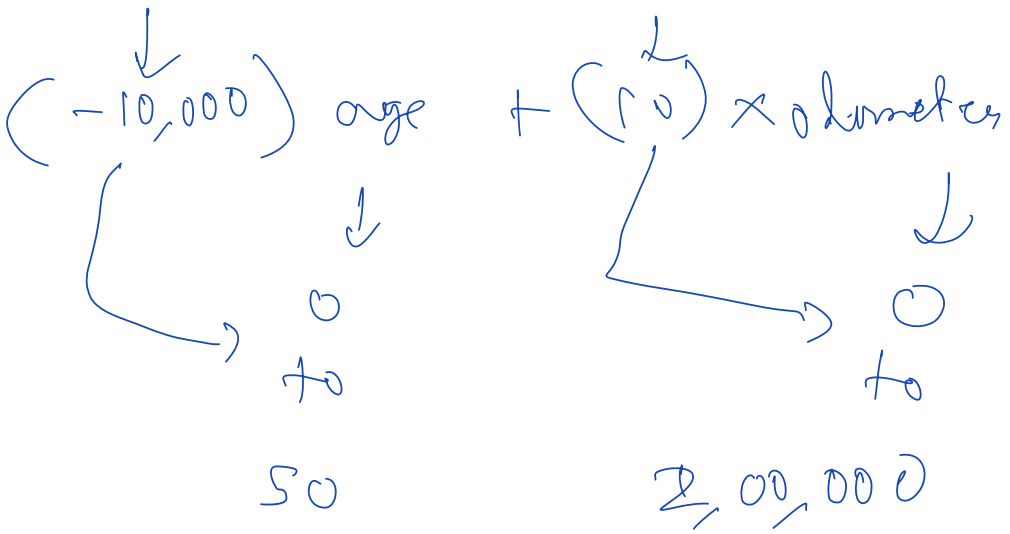
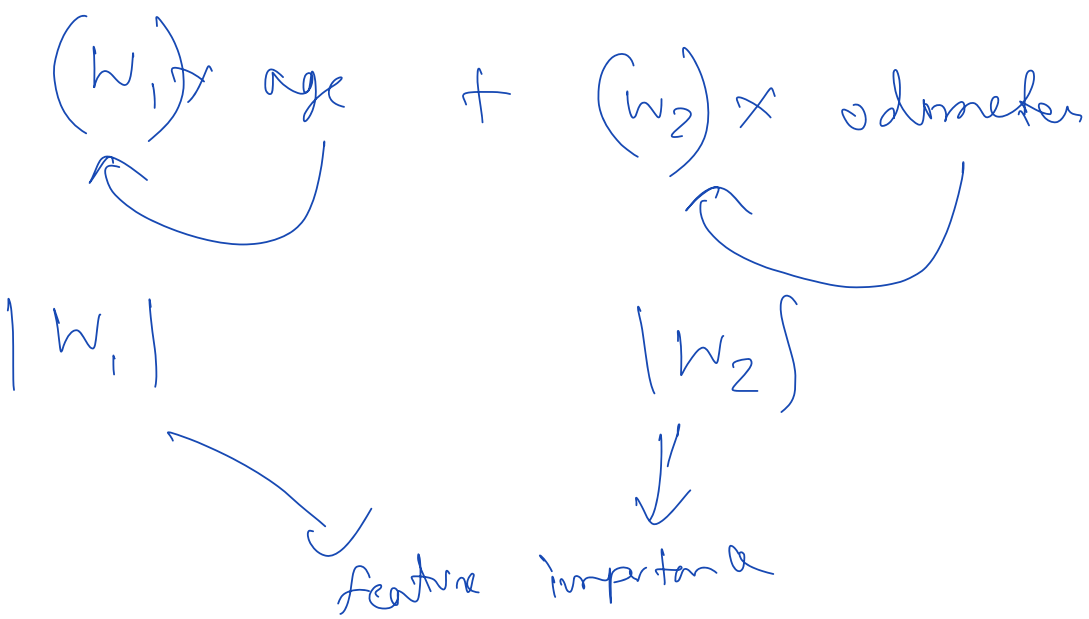
Last class (July 27)

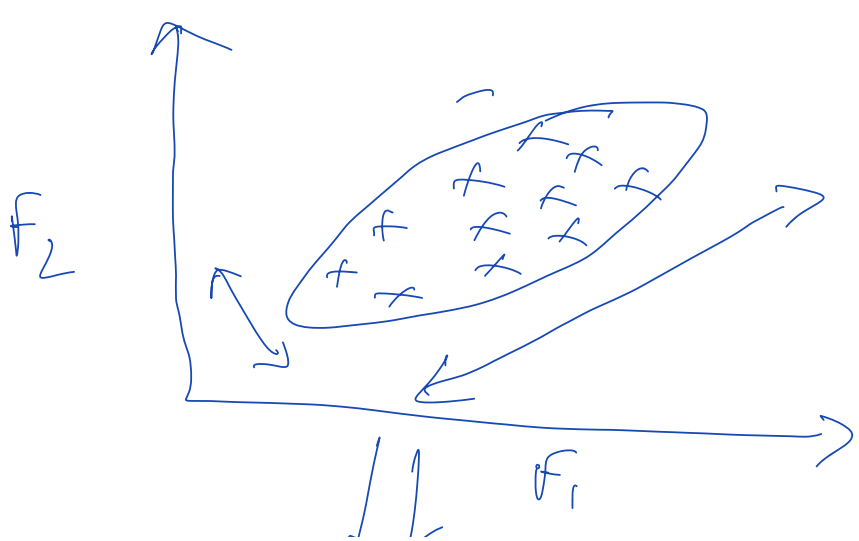
1) Quick recap
2) Target variable encoding
3) Scaling the data
4) Data Notation
5) ML Generalization
6) Lin Reg Intuition
7) Evaluation Metric

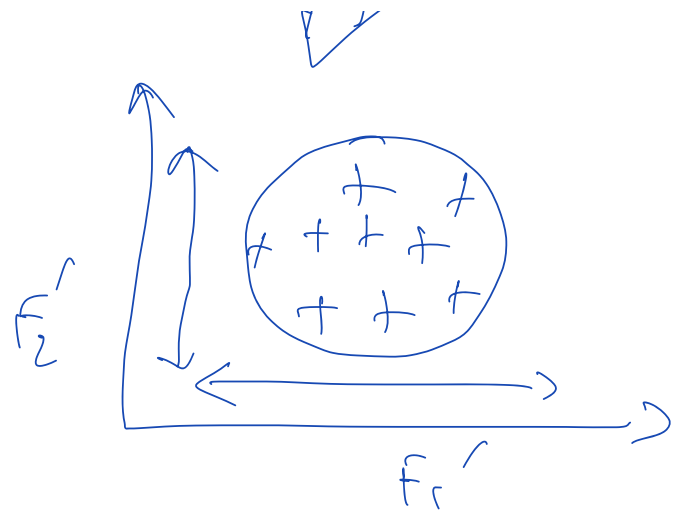
Today's class

1) Quick recap
2) Review of R2 score
3) Model interpretability
4) Revision of Gradient Descent
5) Code for Linear Regression
6) Optimization
7) Implementing Gradient Descent
8) Plot loss function vs weights
9) How feature scaling helps in easier model training.


(-10,000) age
+ (10) x odometer
0
to
50
0
to
2,00,000


(w1) x age
+
(w2) x odometer
|w1|
|w2|
feature importance


F2
F1



Loss = MSE fn

$$= \sum_{i=1}^{m} \left(y^{(i)} - \hat{y}^{(i)} \right)^2$$

error $= \sum_{i=1}^{m} \left(y^{(i)} - \hat{y}^{(i)} \right)$

must differentiate

not necessary to be differentiable

$$\hat{y} = w_0 + w_1 x_1 + w_2 x_2 + \dots w_d x_d$$

$$= b + w^T x$$

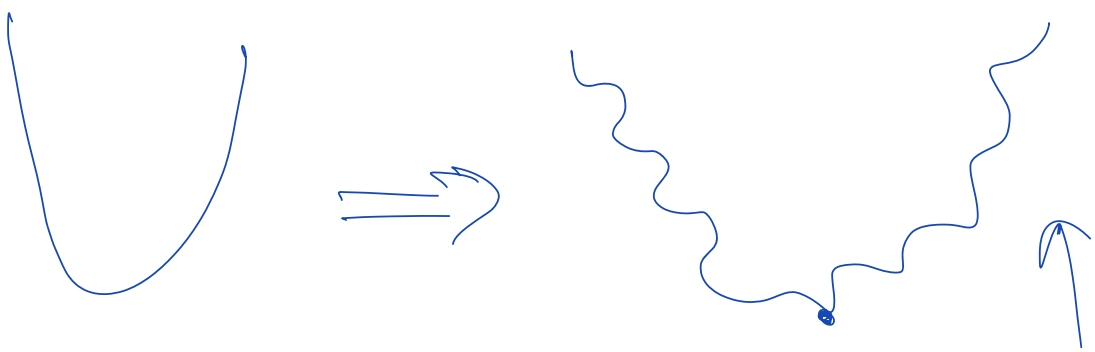
$$w = \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_d \end{bmatrix}, \quad x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_d \end{bmatrix}$$

$(d \times 1)$ $(d \times 1)$

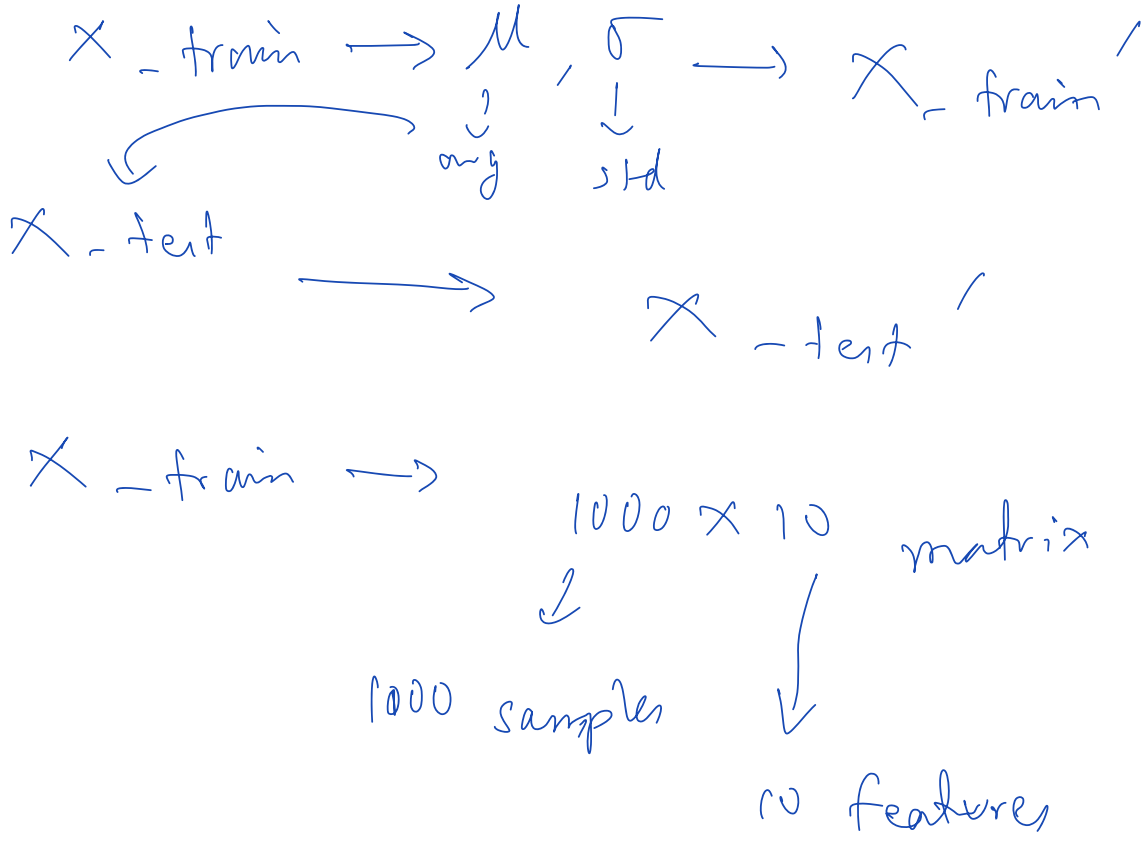
$w^T = 1 \times d$

$x = d \times 1$

$w^T x = (1 \times d) \times (d \times 1) = 1$



10:26 → Resume



$X\_train \rightarrow$ 1000 × 10 matrix

1000 samples

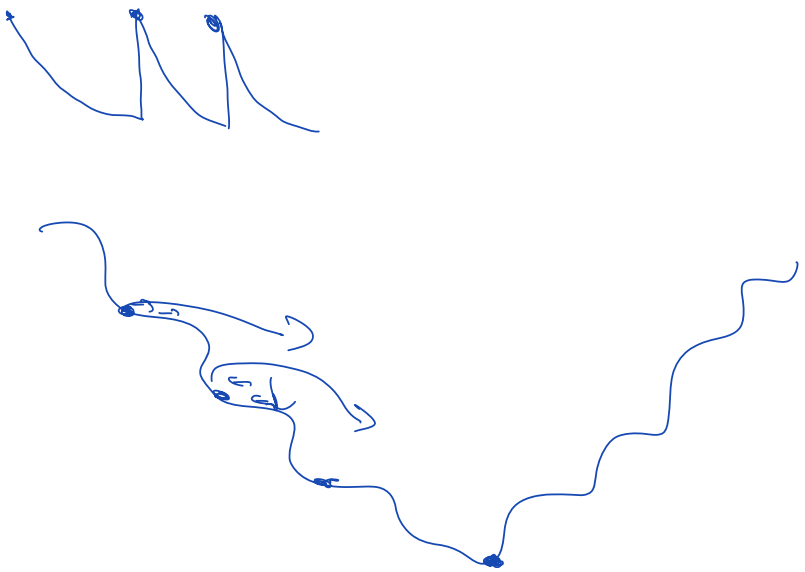
10 features

$\mu \rightarrow 1 \times 10$

$\sigma \rightarrow 1 \times 10$

$$\eta = \frac{k}{i}$$

$k, \frac{k}{2}, \frac{k}{3}, \frac{k}{4}, \ldots$



W → 5.0, 4.999, 4.998, ...

→ (1.0 × 0.1) → 0.1

5.0 → 4.9, 4.8, 4.7, ...

5.0 — 4.999

(0.001) < 0.1

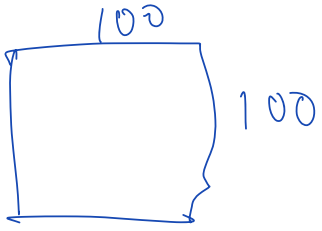
$W = W - 0.1$

SGD: m x d

100 | 10,000

genetic data

$\frac{\partial L}{\partial w_1}$ → 10,000 | 100 x 100



100 x 10000 calculation 10,000 vector

1 ms

10^6 ms → 1000s →

BGD: 100 samples × 10,000 d

↓

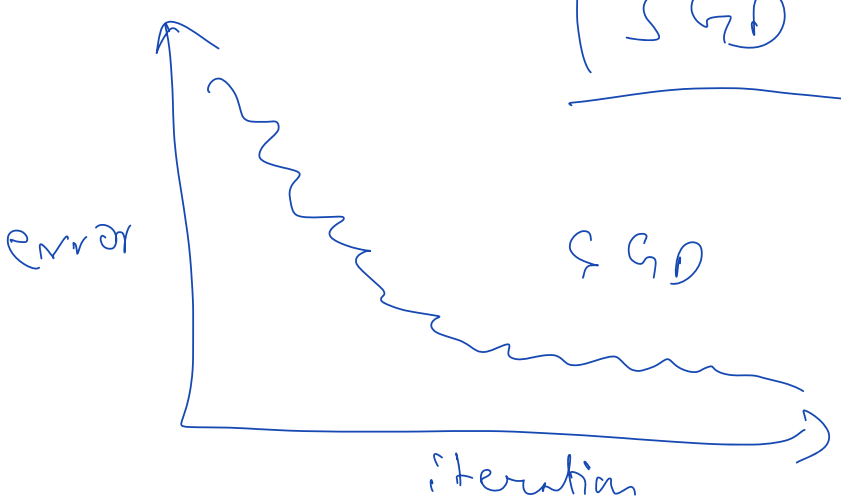
$10^4 \times 5$ ms

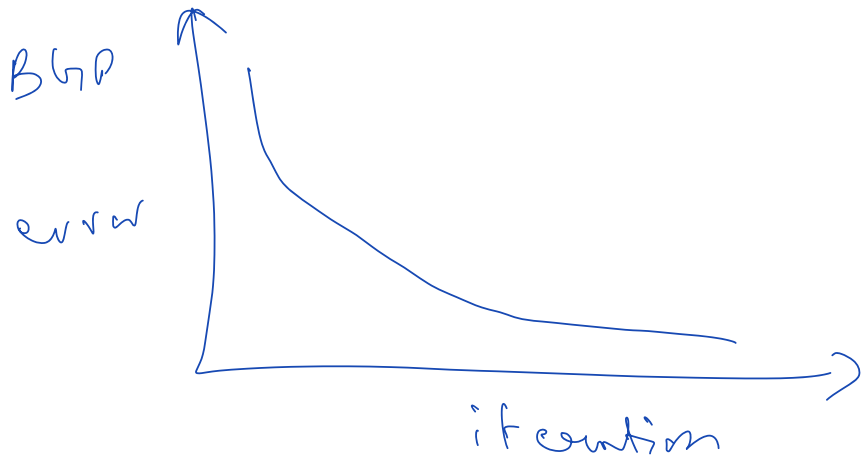
5×10^4 ms → 50 s

10,000 samples × 10 d

↓

SGD ??





1
2
3
→ 4
5
→ 6
7
8 → test
9
10 →

8 → train
2 → test

80/20

random_state = k_2
= 10

test

$$\begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} \quad \begin{bmatrix} x_{11} & x_{21} \cdots x_{d_1} \\ x_{21} & x_{22} \cdots x_{d_2} \\ x_{n_1} & x_{n_2} \cdots x_{d_n} \end{bmatrix} \quad \begin{bmatrix} w_0 \\ w_1 \\ \vdots \\ w_d \end{bmatrix}$$

X

W