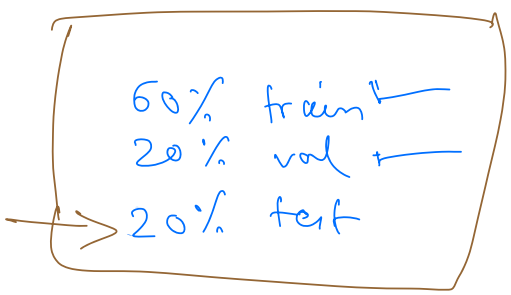
Last class – June 1, 2023

1) Review of Heteroskedasticity
2) Mini Batch & Stochastic Gradient Descent
3) Polynomial Regression

Today – June 3, 2023

1) Recap – Quizzes
2) Generalization & Occam's Razor
3) Underfitting, Overfitting & Tradeoff
4) Bias & Variance
5) Regularization – L2, L1, Elastic Net
6) Hyperparameter tuning wrt Regularization
7) AMA



I) Generalization

Degree of Polynomial

1

should work for both train data & val data

give acceptable error/accuracy

II) Occam's Razor

Pick the simplest model which does the job well

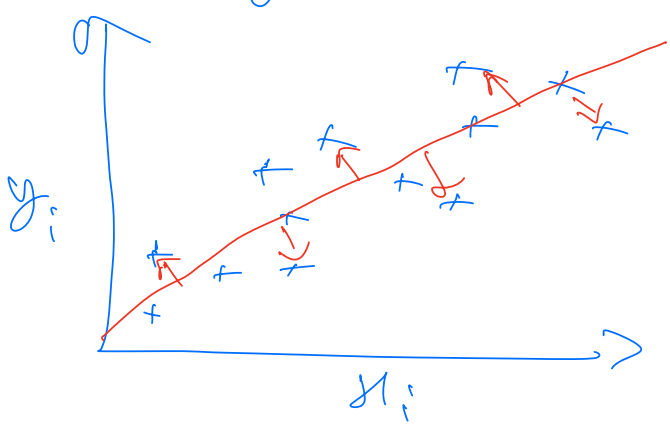
$d = [1, 2, 3, 4, 5]$ → simplest

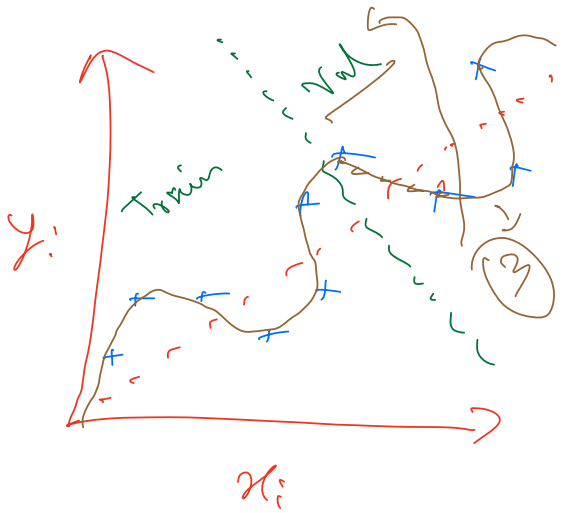
1, 2 ✗

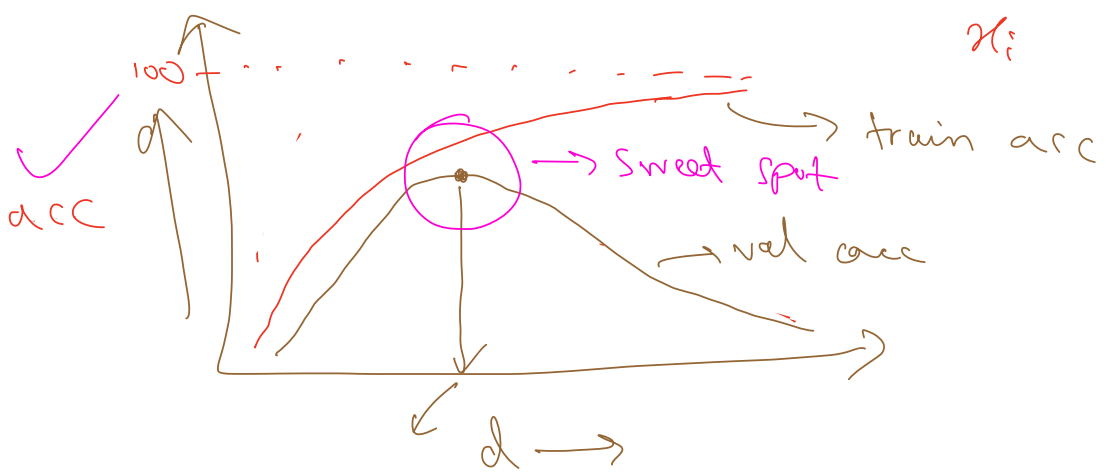
3, 4, 5 ✓

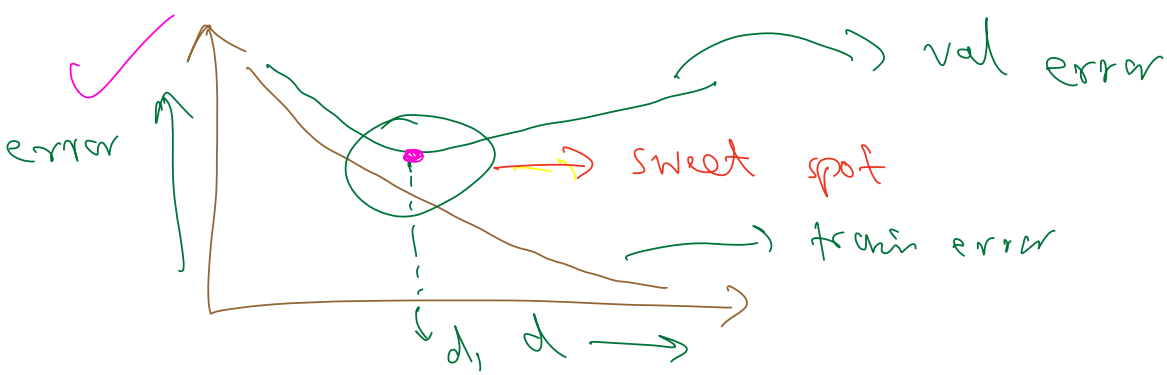
Combining I & II, ⟹ pick the simplest model which gives best accuracy or error on val data

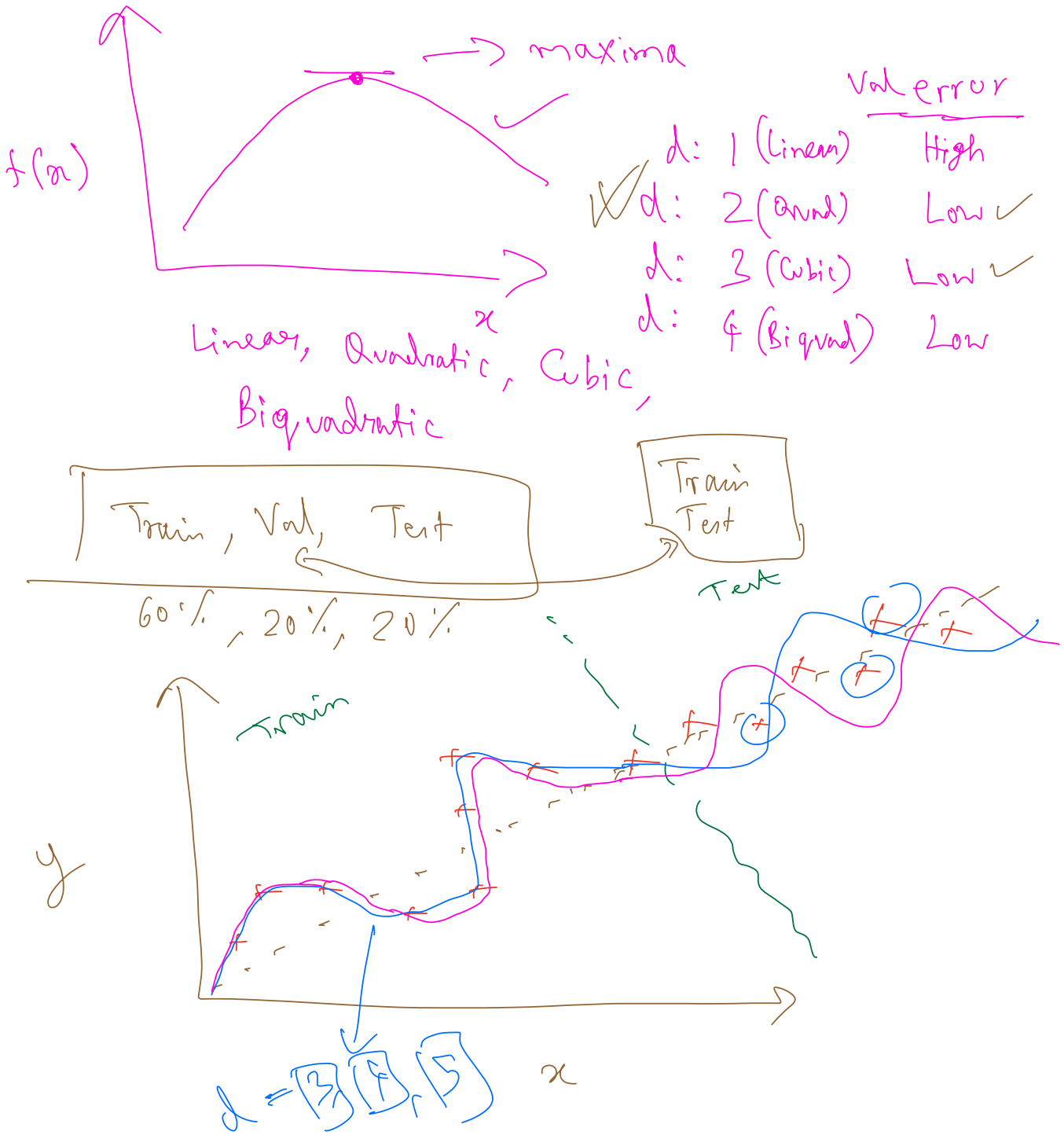
choose d which gives highest R2 score / accuracy / min. error on val data

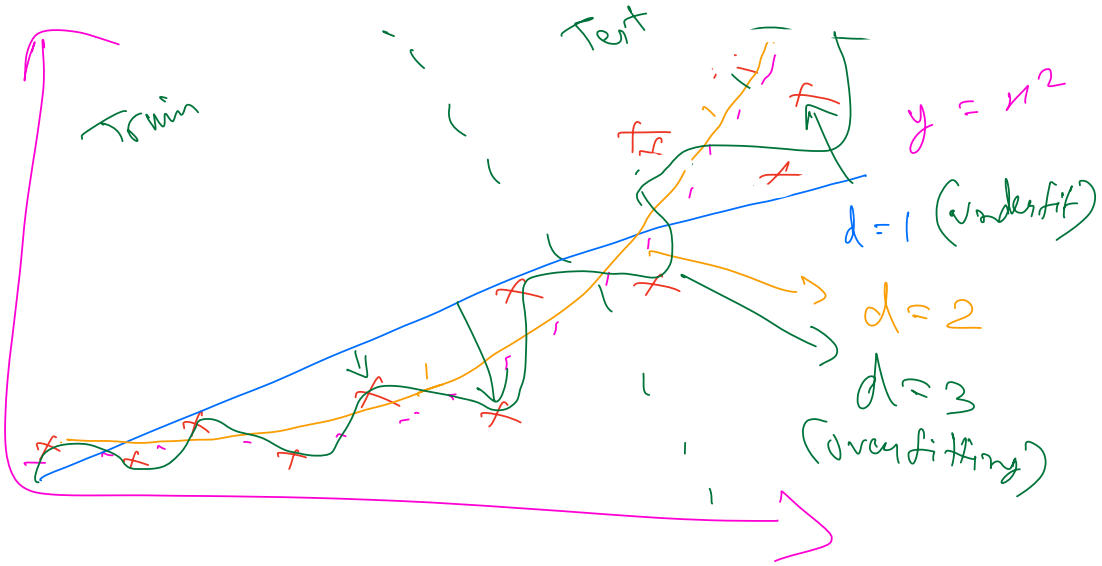









maxima
f(x)
x
Val error
d: 1 (Linear) High
d: 2 (Quad) Low
d: 3 (Cubic) Low
d: 4 (Biquad) Low
Linear, Quadratic, Cubic, Biquadratic
Train, Val, Test
Train Test
60%, 20%, 20%
Test
Train
y
x
d = 3

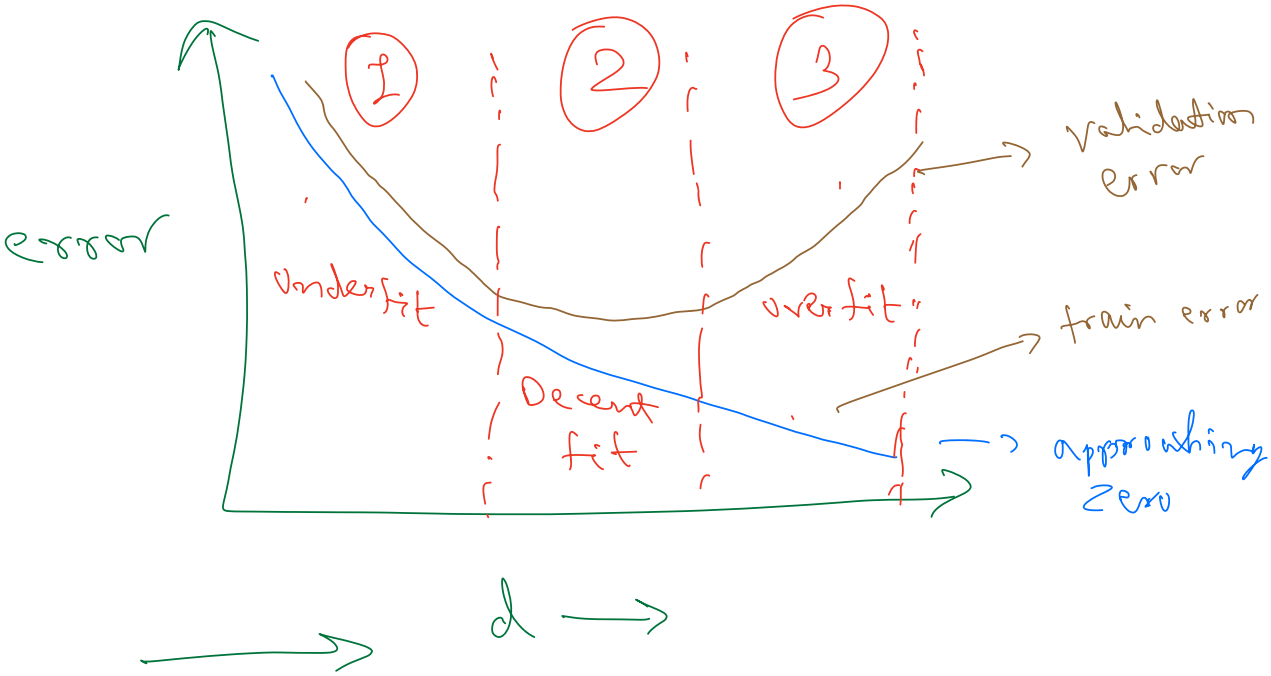


Underfitting, Overfitting & Trade-off

Train error → Test error

low → overfitting ← high

high → underfitting ← high



train error → high, test error → low → very less

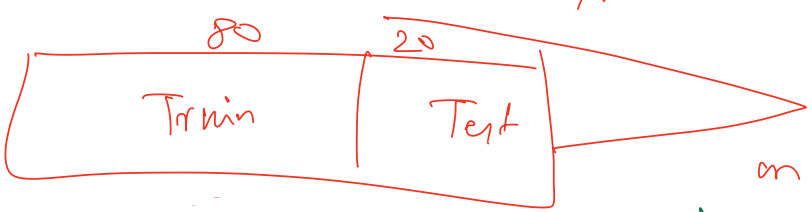
→ Dataset hygiene

→ train/test split is completely random

→ Data leakage

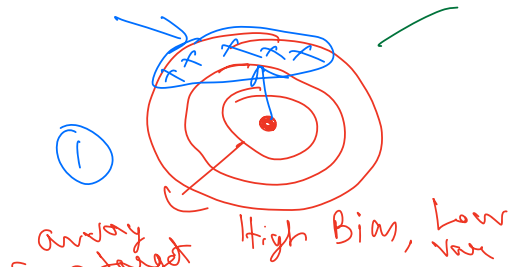
→ test data is known

→ model has been trained on full/partial test data

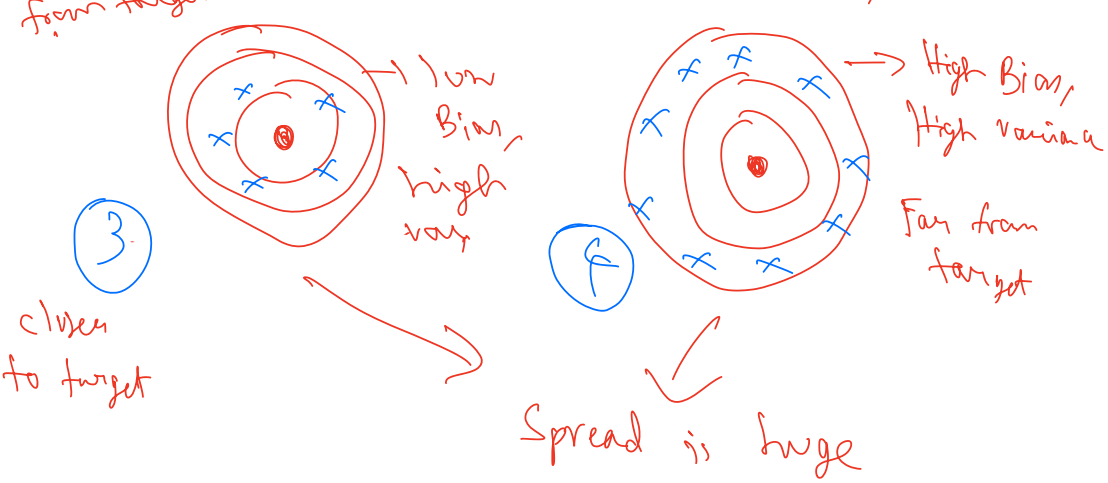


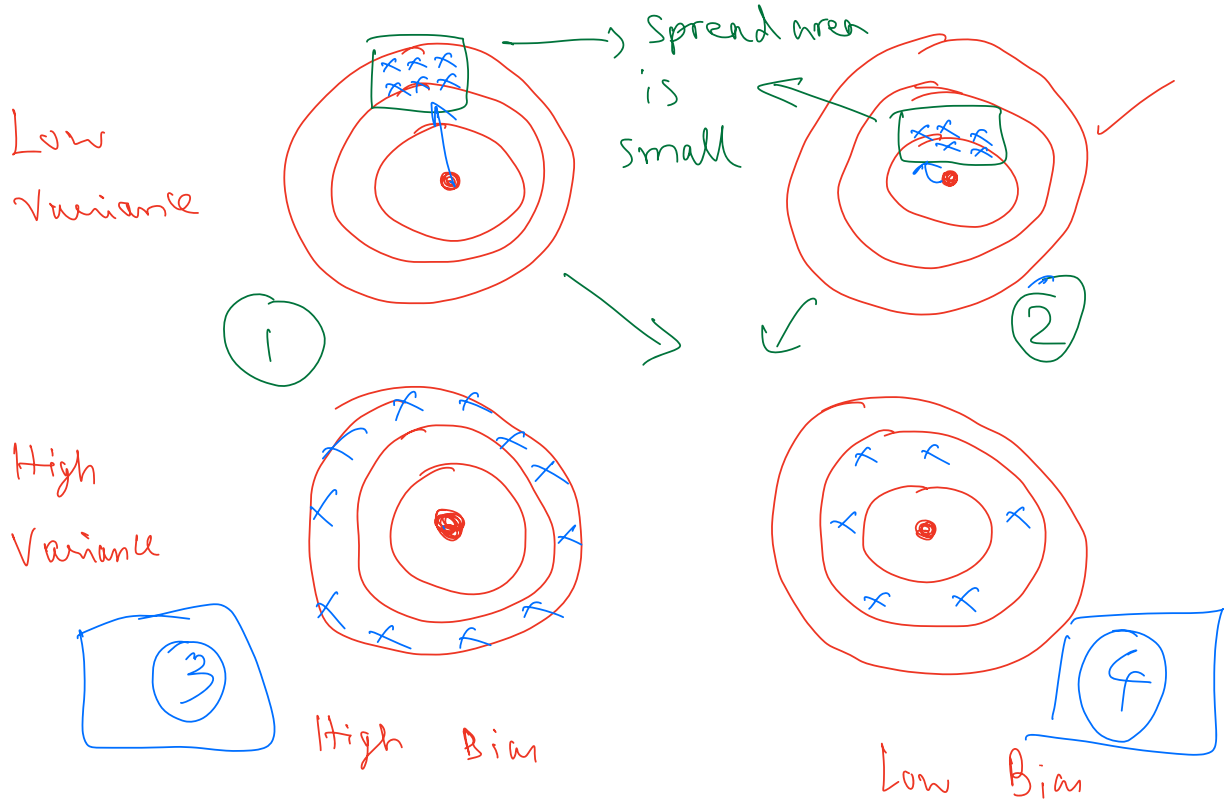
model is trained on whole 80 + 20

Low Variance:



High variance:





1 → Low Var, High Bias

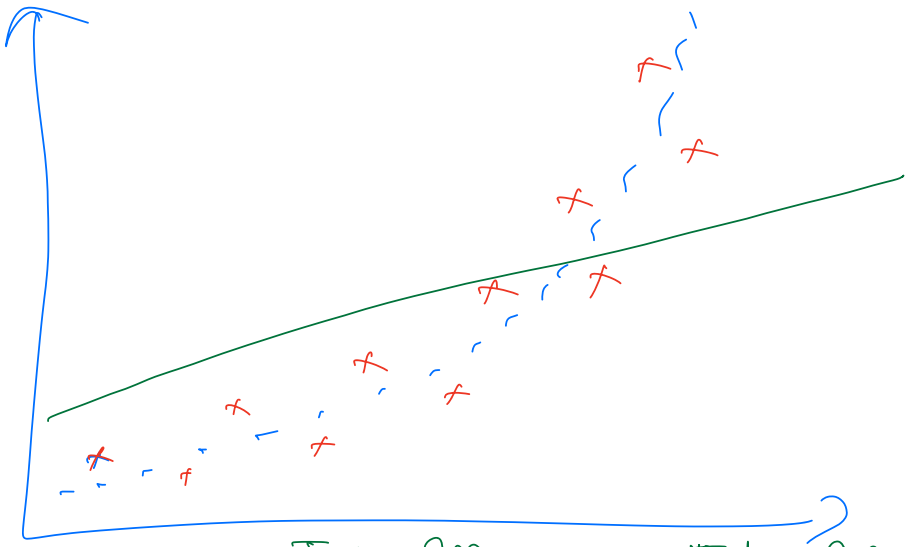
2 → Low Var, Low Bias

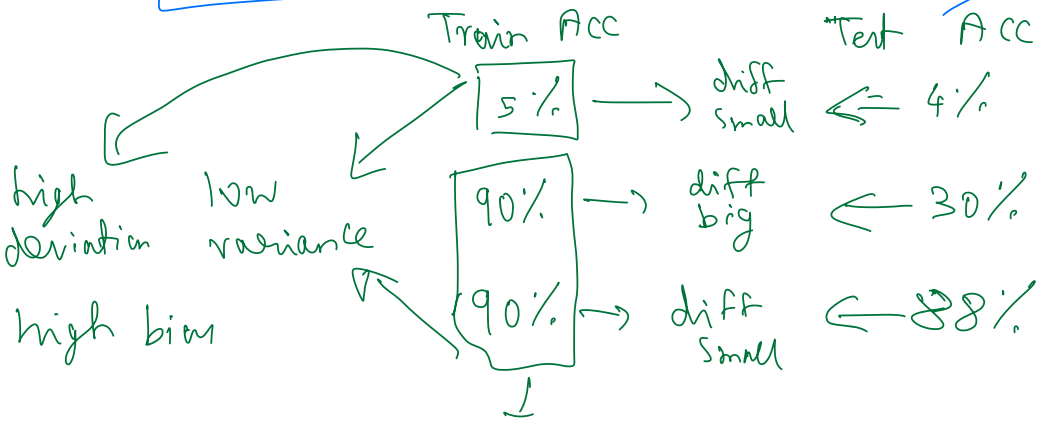
3 → High Var, High Bias

4 → High Var, Low Bias

Underfitting → Train error is high → High Bias
Val error is high

Overfitting → Train error is low → High variance
val error is high





1) If train acc is high, train error is low → low bias
low deviation from target

train acc is low → high bias

2) Diff between train acc & val acc is high → high variance

Diff between train acc & val acc is low → low var

High Bias High Variance

Low → Train acc → 15%] large gap
high bias → Test acc → 2%]

d = 40 → train set] overfitting
d = 5 → test set]

Underfitting → increase the degree of the model (complexity)

overfitting → decrease the degree of the model

$$y = w_0 + w_1 x + w_2 x^2 + w_3 x^3 + w_4 x^4$$

($w_1 \neq 0$, $w_2 \to 0$, $w_3 \to 0$, $w_4 \to 0$)

Loss fn: (Loss to minimized)

$$\sum_{i=1}^{N} \left(y_i - (w_0 + w_1 x_i + w_2 x_i^2 + w_3 x_i^3 + w_4 x_i^4) \right)^2$$

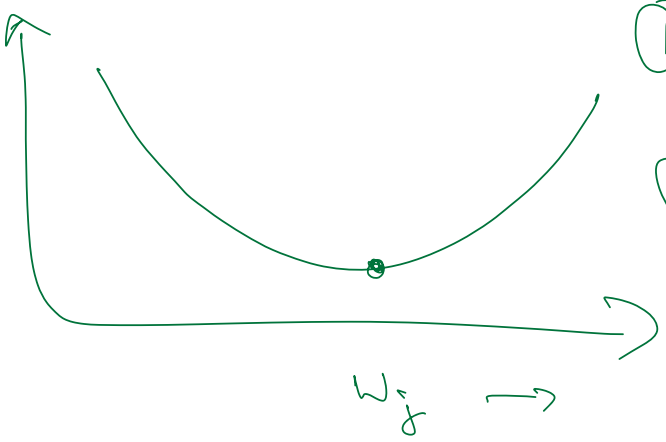
Regularization Term: $+ \lambda (w_1^2 + w_2^2 + w_3^2 + w_4^2)$

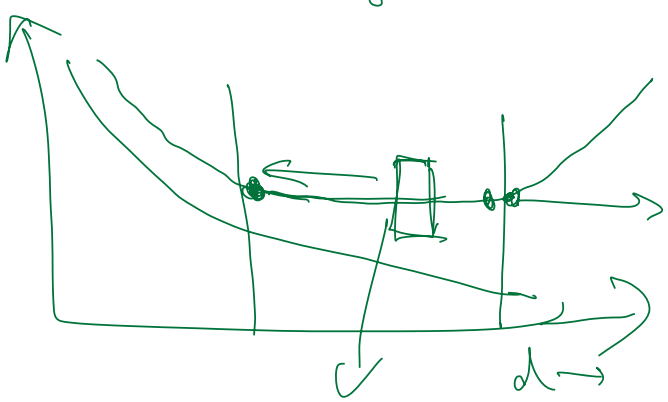
→ ⟷

$$\min \quad \lambda \left(w_1^2 + w_2^2 + w_3^2 + w_4^2 \right)$$

(1) $\left(y_i - \sum_j w_j x_{ij} \right)^2$

(2) $\lambda \sum_j w_j^2$





Regularization

w_1 (salary)

↓

1 L to 10 L

0 to 1

$\bar{w}_1 \rightarrow$ low

w_2 (kids)

2 to 4

0 to 1

$\bar{w}_2 \rightarrow$ high

$$\begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1,10} \\ & & & \\ x_{N1} & x_{N2} & \cdots & x_{N,10} \end{bmatrix}_{N \times 10} \begin{bmatrix} w_1 \\ w_2 \\ w_3 \\ \vdots \\ w_{10} \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix}_{N \times 1}$$

$$X \quad [W] = Y$$

$x_1, x_2, x_3 \ldots x_{1,10}$

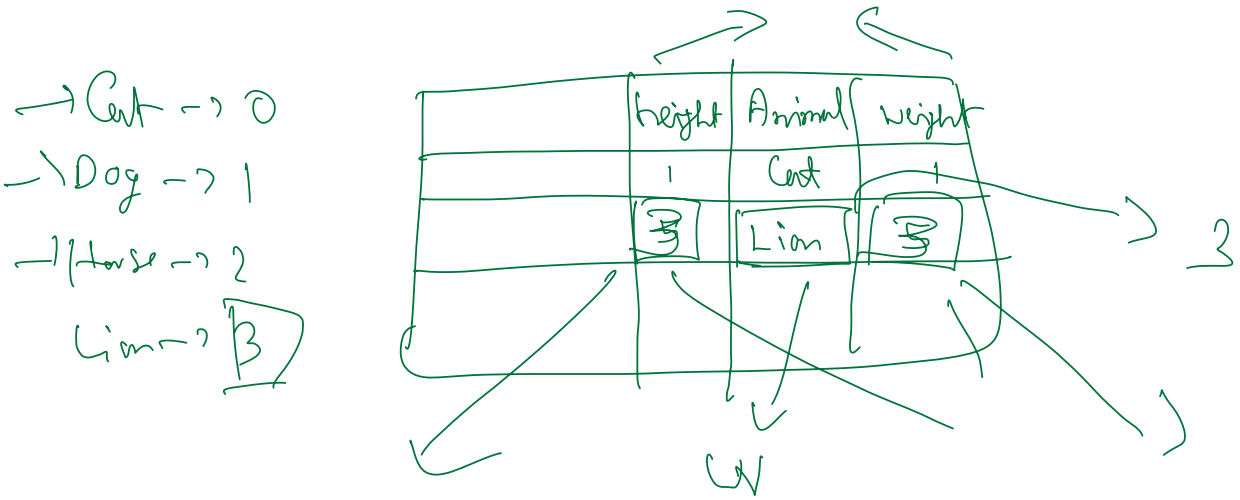
$x_1, x_2, x_3 \longrightarrow$

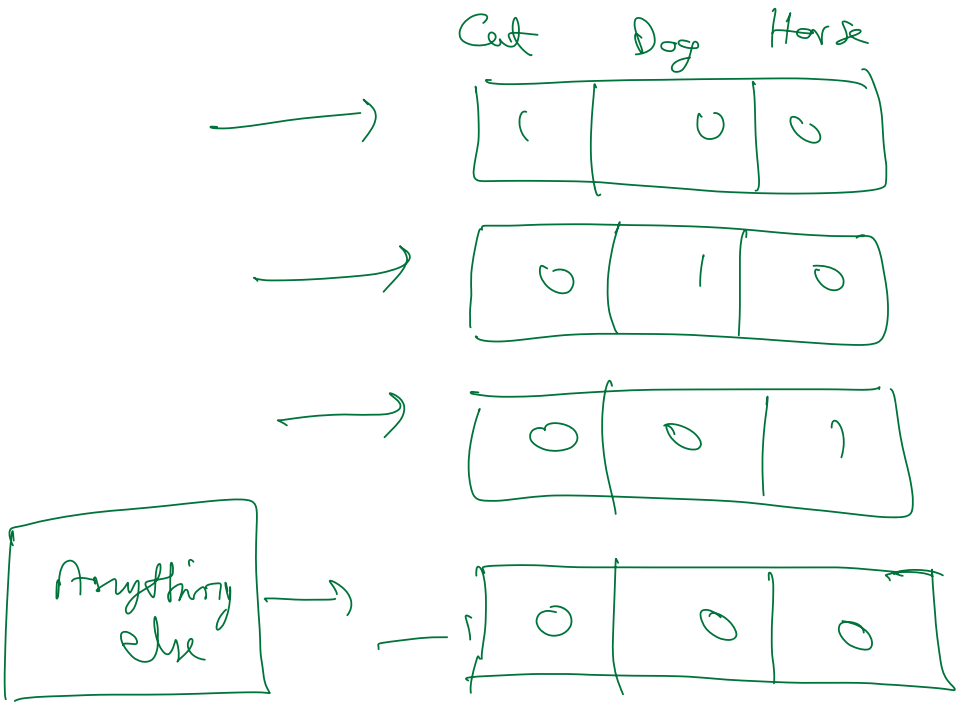
$x_1, x_2, x_3, x_1^2, x_2^2, x_3^2, x_1x_2, x_1x_3, x_2x_3$

Collinear $x_2 = \sqrt{x_1}$ $\quad x_2^2 \rightarrow x_1$

	Height	Animal	Weight
	1	Cat	1
	2	Dog	2
	~~3~~	Horse	~~3~~

Training

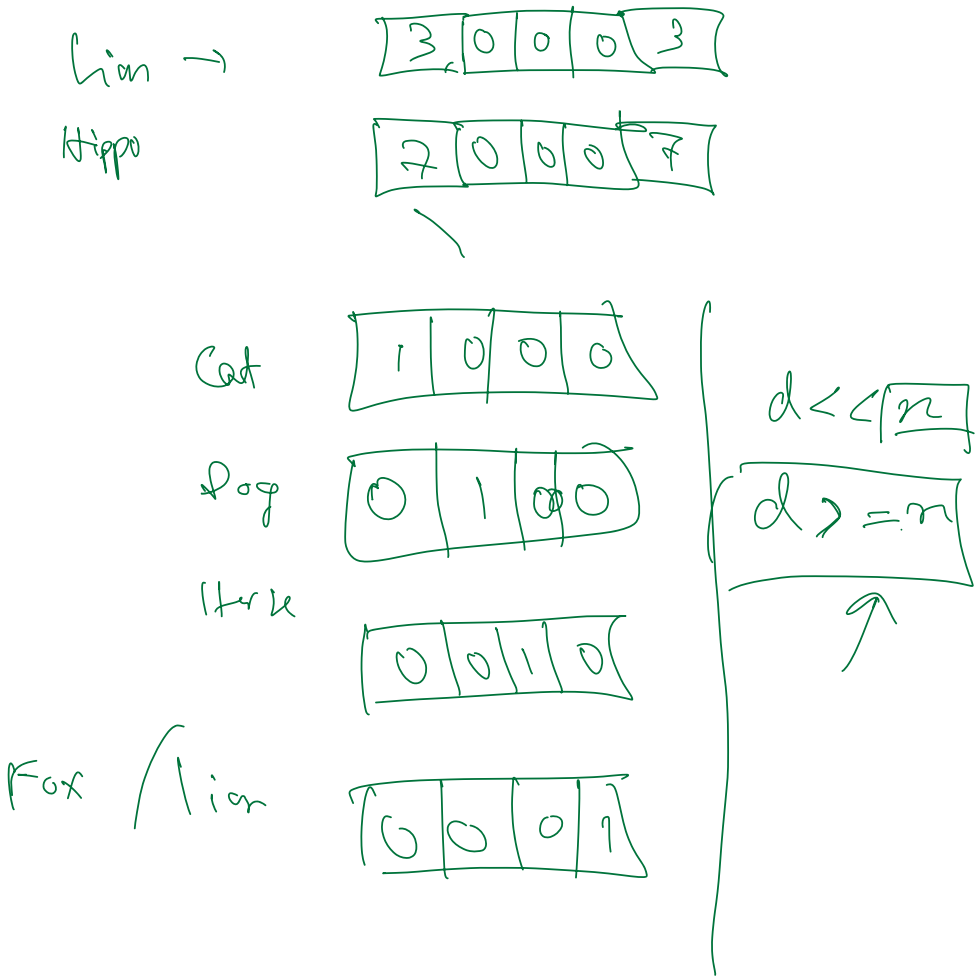




N ~~dau~~ Categories

N-1 columns

Predict the food needed for the animal


Lion
3 0 0 0 3
Hippo
2 0 0 0 2
Cat
1 0 0 0
Dog
0 1 0 0
Horse
0 0 1 0
Fox
0 0 0 1
d << n
d >= n