

Laplace Smoothing

span / hnn

$N=100$ emails

50 span / 50 hnn

$$P(A|B) = \frac{P(B|A) P(A)}{P(B)}$$

$$P(S=1 | W_i) = \frac{P(W_i | S=1) P(S=1)}{P(W_i)}$$

'win one million dollars'

$$P(\text{win, one} \dots | S=1)$$

$$= P(\text{win} | S=0) P(\text{one} | S=0)$$

$$\cdot \cdot P(\text{dollars} | S=0)$$

$$P(\text{win} | S=0) = \frac{20}{1000} P(\text{one} | S=0) = \frac{1}{1000}$$

$$P(\text{dollars} | S=0) = 0$$

$$P(w_i | S=0) = P(\text{dollars} | w_i | S=0) = 0$$

$$P(S=0 | w_i) = 0$$

MLE fails

$$P(\text{dollars} | S=0) = \frac{0+1}{1000+2}$$

$$= \frac{1}{1002}$$

$k=1$ Laplace

$$\frac{\text{counts} + K}{\text{total counts} + K \cdot C}$$

$$P(\text{win} | s=0) = \frac{20}{1000} \Rightarrow \frac{20+1}{1002}$$

$$= \frac{21}{1002}$$

$$K=0.5$$

$$P(\text{win} | s=0) = \frac{20+0.5}{1000 + (0.5)K}$$

$$= \frac{20.5}{1001}$$

Bias-Variance Trade-off

$$E(\underline{y} - \underline{\hat{y}}) = \text{Var} + \text{Bias}$$

$$\underline{\text{Var}} = \text{MSE}_{\text{Test}}$$

$$\underline{\text{Bias}} = \underline{\text{MSE}_{\text{Training}}}$$

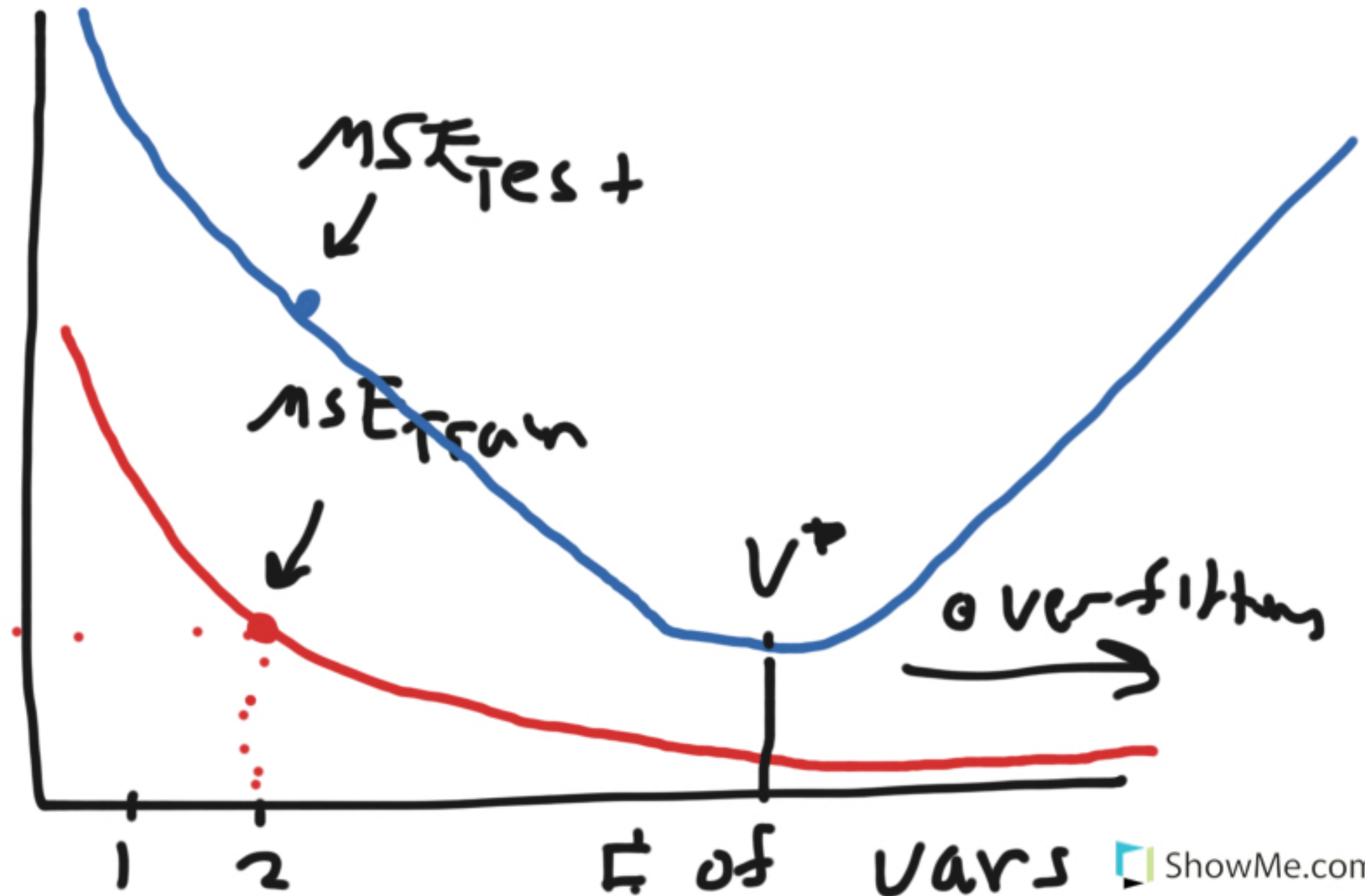
As Complexity $\uparrow \Rightarrow \text{MSE}_{\text{Training}} \downarrow$
" $\uparrow \Rightarrow \text{MSE}_{\text{Test}} \uparrow$ or \downarrow

$$C_i = \alpha + \beta_1 \text{pop density} +$$

$C_i \equiv \text{casualties}$

$\beta_2 \text{ wcap}$

MSE



Decision Trees

o divide and conquer \Rightarrow break up
the data \rightarrow analyze subsets

Can we predict whether an attack will be an 'assassination' (assn) or bombing (bomb) feature?

Region	Day	Weapon	Type
1	1	Fir	Assn
2	2	Bom	Assn
1	3	Fir	Bomb
1	1	Fir	Assn
2	2	Fir	Assn
1	1	Bom	Bomb
2	20	Bom	Assn

5/2
Region

1 2
2/2 (3/0)
Day Assn

1 3
2/1 0/1

5 (Assn) / 2 (Bomb)

Resian

~~(2/2)~~ ①

✓ ② (3/0) Assn

Day

New Infi

Resian = 1
Day = 1
Weapon = Bomb
Type ?

~~(2/1)~~ ①

✓ ③ (0/1) Bomb

Weapon

Fire

✓ ④ (2/0) Assn

Bomb

✓ ⑤ (0/1) Bomb

Decision tree algo

- ① choose the best splitting feature based on info gain
- ② create a new child node
- ③ for each child node:
if (subset is pure) {stop}
else split (node, examples)



① They predict very well
(high F_1 , Accuracy etc)

② Transferred

Random Forests

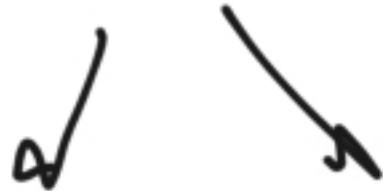
1 $\begin{cases} p = 1 \\ n = 1 \\ m = \text{Bomb} \end{cases}$

2 trees

① Region

②

Day

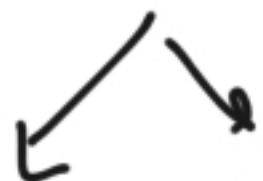


①

②

2/1

1/0



Fir Bomb

Bomb

1

2

3

20

warm

③

Assn

Assn

2/3 vote

Assn