$$w_1 (x^{(1)})^2 + w_2 \cdot |x^{(1)}| + w_3 (x^{(1)} + x^{(2)}) = H(x^{(1)}, x^{(2)})$$

← form of eqn changes matrix

$x^{(1)}$	$x^{(2)}$	y
1	2	5
3	6	0
4	1	1

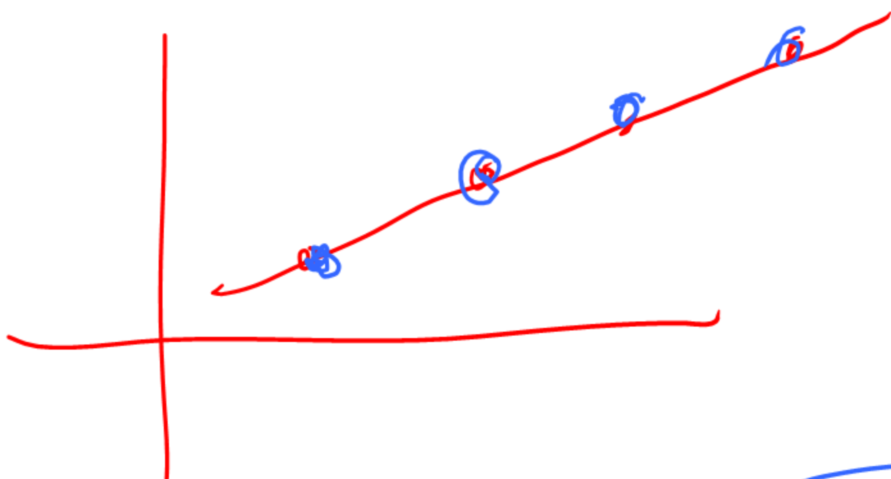
$$\underbrace{\begin{bmatrix} 1 & 1 & 3 \\ 9 & 3 & 9 \\ 16 & 4 & 5 \end{bmatrix}}_{X} \underbrace{\begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix}}_{w} \qquad \underbrace{\begin{bmatrix} 5 \\ 0 \\ 1 \end{bmatrix}}_{y}$$

$$\begin{bmatrix} H(x^{(1)}_1, x^{(2)}_1) \\ H(x^{(1)}_2, x^{(2)}_2) \\ H(x^{(1)}_3, x^{(2)}_3) \end{bmatrix}$$

← Xw is predictions

MSE is how different Xw are from y

ideally, we'd have $Xw = y$



typically $X\vec{w} \approx \vec{y}$

$$\vec{0} \approx \boxed{X\vec{w} - \vec{y}}$$

$$\|\vec{v}\|^2 = \left(\sqrt{v_1^2 + v_2^2 + \dots + v_{\#comp}^2}\right)^2$$

$$\left((Xw)_1 - y_1\right)^2 + \left((Xw)_2 - y_2\right)^2 + \dots = \|X\vec{w} - y\|^2$$

the theory says:

the way to minimize MSE is by solving normal eqns.

$$\boxed{X^T X w = X^T y}$$

$$X^T X w = X^T y$$

parameters

let's pretend $w_1 = 7$, $w_2 = -3$,

$$w_3 = 5$$

Solves normal eqns

this means

$$7(x^{(1)})^2 - 3 \cdot |x^{(1)}| + 5(x^{(1)} + x^{(2)})$$

has smaller MSE than any other eqn of the form

$$w_1 (x^{(1)})^2 + \overset{-3}{w_2} |x^{(1)}| + \cancel{\overset{0}{w_3} (x^{(1)} + x^{(2)})}$$

which has smaller MSE, the sol'n to the normal eqns where

$$X = \begin{array}{c} \\ \\ \end{array}\begin{array}{ccc} & x^{(1)} & x^{(2)} \\ \end{array}$$

$$X = \begin{bmatrix} 1 & 3 & 5 \\ 1 & 8 & 6 \\ 1 & 6 & 9 \\ 1 & 2 & 0 \end{bmatrix}$$

or

plane

$$H(x^{(1)}, x^{(2)}) = w + w_1 x^{(1)} + w_2 x^{(2)}$$

linear combo

$$X = \begin{array}{c} \\ \end{array}\begin{array}{ccc} & & x^{(2)} \end{array}$$

$$X = \begin{bmatrix} 1 & 3 & 4 \\ 1 & 8 & 9 \\ 1 & 6 & 7 \\ 1 & 2 & 3 \end{bmatrix}$$

$$H(x^{(1)}, x^{(2)}) = w_0 + w_1 x^{(1)} + w_2 x^{(2)}$$

$$\text{where } x^{(2)} = \underbrace{x^{(1)} + 1}_{\text{plug in for } x^{(2)}}$$

$$H(x^{(1)}, x^{(2)}) = w_0 + w_1 x^{(1)} + w_2 (x^{(1)} + 1)$$

line

$y = w_0 + w_1 x$

parabola

$y = w_0 + w_1 x + w_2 x^2$

$= \underline{w_0} + w_1 x^{(1)} + w_2 x^{(1)} + \underline{w_2}$

$= \underbrace{(w_0 + w_2)}_{\text{line}} + \underbrace{(w_1 + w_2)}_{} x^{(1)}$

$= \underline{c_0 + c_1 x^{(1)}}$

Counting/Prob.

52 cards, 13 in each of 4 suits (♥, ♦, ♠, ♣)

someone deals you a hand of 5 random cards

$P(\text{all ♥s}) = \frac{13}{52} \times \frac{12}{51} \times \frac{11}{50} \times \frac{10}{49} \times \frac{9}{48}$

$$= \frac{P(13,5)}{P(52,5)}$$ sequences of 5 cards

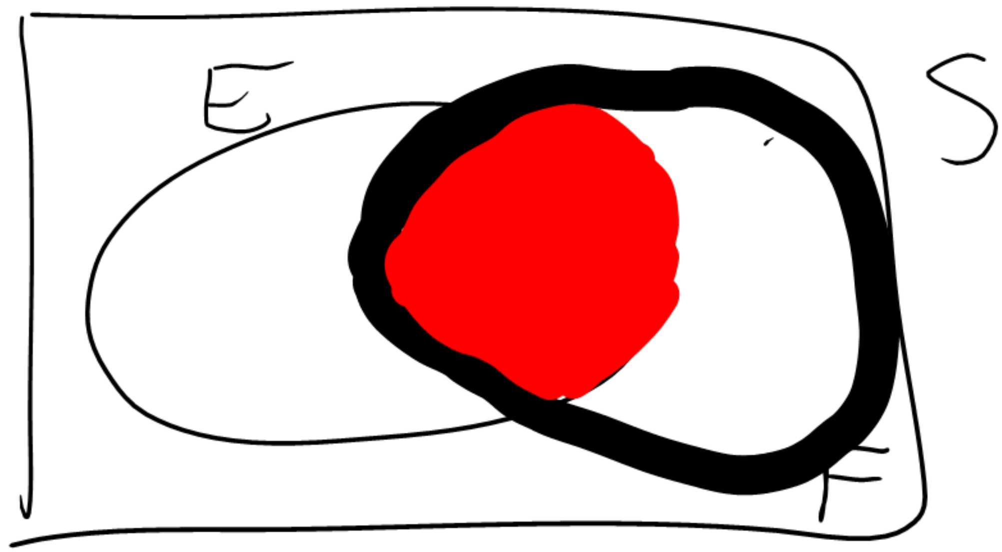
$$= \frac{C(13,5)}{C(52,5)}$$ sets of 5 cards

P(all 5 cards have distinct values) $= \underbrace{1}_{\text{say } K} * \underbrace{\frac{48}{51}}_{\text{not } K, \text{ say } 7} * \underbrace{\frac{44}{50}}_{\text{not } K, \text{ not } 7} * \ldots,$

$P(\text{all 5 cards have distinct values} \mid \text{all } \heartsuit\text{s})$

$= 1$

$$P(E|F) = \frac{P(E \cap F)}{P(F)}$$



$$P(\text{no} \heartsuit) = \frac{C(39, 5)}{C(52, 5)}$$

sets of 5 cards

$$P(\text{all} \heartsuit) \neq 1 - P(\text{no} \heartsuit)$$

$$P(\text{at least } 1 \heartsuit) = 1 - P(\text{no} \heartsuit)$$

$$P((\text{all } \heartsuit) \cup (\text{exactly } 1\ K))$$

$$\text{Prob}(\text{all } \heartsuit) + \text{Prob}\begin{pmatrix}\text{exactly}\\ 1\ K\end{pmatrix} - \text{Prob}\begin{pmatrix}\text{all } \heartsuit\\ \text{and}\\ \text{exactly}\\ \text{one}\\ K\end{pmatrix}$$

$$\frac{P(13,5)}{P(52,5)} + \frac{\overset{\text{pos-}}{5}\cdot \overset{\text{suit}}{4}\cdot P(48,4)}{P(52,5)} - \frac{5 * P(12,4)}{P(52,5)}$$

non-K _ _ K _ non-K _ _

_ $K\heartsuit$ _ non K must be $\heartsuit$