315 Lecture 3

Announcements

- HW1
  - Due yesterday
- HW2
  - Out soon!
  - Due in two weeks

Website

- Lec2 notes and code
- Topics
- Reading
- Homework due dates

Last Week

- ML Concepts

Today

- Linear Regression

Next Week

- Probabilistic view of regression

linear regression

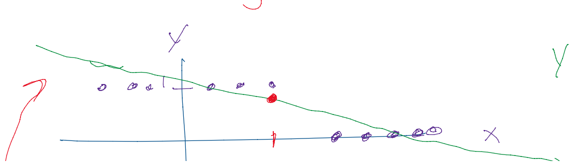
	2-class view	Multiclass
Classification	Drought-resistant	Count kernels
Regression	Prob(Drought-res)	Expected # kernels

$\mathcal{D} = \{(y^{(n)}, x^{(n)})\}_{n=1}^{N}$ $\quad \vec{x}^{(n)} \in \mathbb{R}^M$

$y^{(n)} \in \{0,1\}$

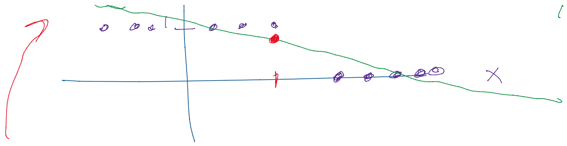
$y \in \{0,1\}$	$y \in \{1,\dots,K\}$
$y \in \mathbb{R}\ [0,1]$	$y \in \mathbb{R}$

Poll 1 Can we solve classification problems with regression?

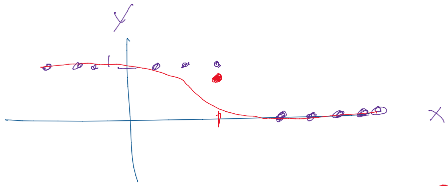


$y = mx + b$

$y > 0.5$



$y > 0.5$ 1
o.w   0



$$g(z) = \frac{1}{1+e^{-z}}$$

[Yes, but we need decision function] $z = mx + b$

Linear

set of points $\{z_i\}$ combine by scaling with constant a_i and summing

$$\sum z_i a_i$$

Poll 2 Is this linear?

$$y = w_1 x_1 + w_2 x_2 + b$$

[No, technically this is affine]

Affine

$$\vec{y} = \underbrace{\vec{w}^T \vec{x}}_{\text{linear}} + b$$

Convert Affine to Linear?

$$X \in \mathbb{R}^{N\times 2} \qquad\qquad X \in \mathbb{R}^{N\times 3}$$

$$\vec{w}^T \; [w_1\ w_2] \begin{bmatrix} x_1^{(1)} & x_2^{(1)} \\ \vdots & \vdots \\ x_1^{(N)} & x_2^{(N)} \end{bmatrix} + \overset{b}{\square} \Longrightarrow [b\ w_1\ w_2] \begin{bmatrix} 1 & x_1^{(1)} & x_2^{(1)} \\ \vdots & \vdots & \vdots \\ 1 & x_1^{(N)} & x_2^{(N)} \end{bmatrix}$$

Whoops, this is broken because the matrix/vector

Whoops, this is broken because the matrix/vector dimensions don't match! Corrected below ↓

$$X \in \mathbb{R}^{N\times 2} \qquad X \in \mathbb{R}^{N\times 3} \qquad \hat{y} \in \mathbb{R}^{N\times 1}$$

$$\begin{bmatrix} x_1^{(1)} & x_2^{(1)} \\ \vdots & \vdots \\ x_1^{(N)} & x_2^{(N)} \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} + b \Longrightarrow \begin{bmatrix} 1 & x_1^{(1)} & x_2^{(1)} \\ 1 & \vdots & \vdots \\ 1 & x_1^{(N)} & x_2^{(N)} \end{bmatrix} \begin{bmatrix} b \\ w_1 \\ w_2 \end{bmatrix} = \begin{bmatrix} b + w_1 x_1^{(1)} + w_2 x_2^{(1)} \\ \vdots \end{bmatrix}$$

$$y = Xw \cancel{+b}$$

Poll 3: Is this linear

$$y = w_1 x + w_2 x^2 + w_3 x^3$$

[Yes, it is linear in the weights]

Assume data is constant; even x^2 and x^3...

$$w_1 3 + w_2 9 + w_3 27$$

$$\vec{x} \in \mathbb{R}^{N\times 1} \longrightarrow \Phi(\vec{x}) \in \mathbb{R}^{N\times 4}$$

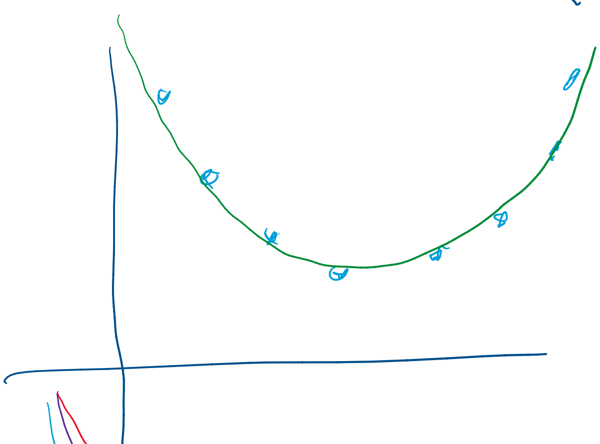
$$\begin{bmatrix} x^{(1)} \\ \vdots \\ x^{(N)} \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & x^{(1)} & x^{2\,(1)} & x^{3\,(1)} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x^{(N)} & x^{2\,(N)} & x^{3\,(N)} \end{bmatrix}$$

$$\vec{y} = X\vec{w}$$

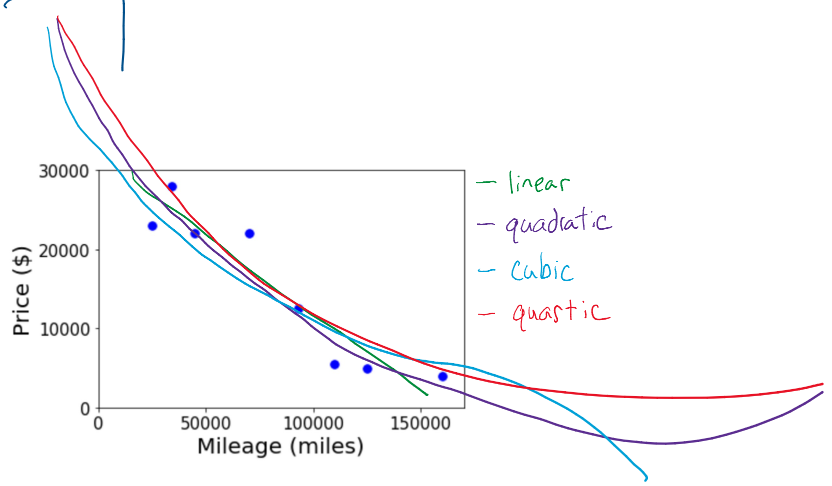
$$\left[\quad\right] = \left[\quad\quad\right] \begin{bmatrix} w_0 \\ w_1 \\ w_2 \\ w_3 \end{bmatrix}$$

$$\vec{x}_{new} = \phi(x) = \begin{bmatrix} 1 \\ x \\ x^2 \\ x^3 \end{bmatrix}$$

"feature" or "basis" function



$w_0 + w_1 x + w_2 x^2$



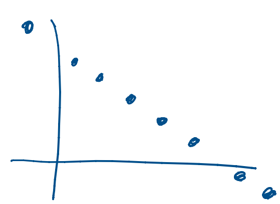
$y = Xw$

Poll 4: $w = [w_0\ w_1\ w_2\ w_3]^T$

$y = w_0 + w_1 x_1 + w_2 x^2 + w_3 x^3$

Can I learn a line hypothesis

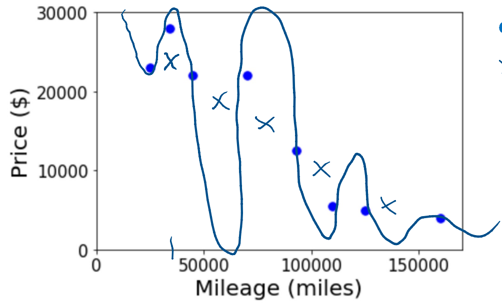
$y = w_0 + w_1 x_1$



[Yes, it would need to learn that

[Yes, it would need to learn that $w_2 = 0$ and $w_3 = 0$]

$y = w_0 \; \cdots\cdots \; w_{100} x^{100}$



Overfitting!
Model complexity is too high

$$J(\vec{w}) = \frac{1}{2}\frac{1}{N}\|\vec{y} - X\vec{w}\|_2^2 \longrightarrow \vec{w} = (X^T X)^{-1} X^T \vec{y}$$

Two general ways to min $J(\vec{w})$
obj func

1) Gradient Descent
   0) Start with rand $\vec{w}$
   1) Repeat
      for all $i \in \{1, .., M\}$

$$w_i \leftarrow w_i + \alpha \frac{\partial J}{\partial w_i}$$

$$\vec{w} \leftarrow \vec{w} + \alpha \nabla_{\vec{w}} J(\vec{w})$$

2) Closed form Solution
   1) $\nabla J(w) = 0$
   2) solving for w

- Expand objective
  - $J(\boldsymbol{w}) = \frac{1}{2} \|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{w}\|_2^2$
  - $= \frac{1}{2} (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{w})^T (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{w})$
  - $= \frac{1}{2} \left(\boldsymbol{y}^T - \boldsymbol{w}^T \boldsymbol{X}^T\right)(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{w})$
  - $= \frac{1}{2} [\boldsymbol{y}^T \boldsymbol{y} - \boldsymbol{w}^T \boldsymbol{X}^T \boldsymbol{y} - \boldsymbol{y}^T \boldsymbol{X}\boldsymbol{w} + \boldsymbol{w}^T \boldsymbol{X}^T \boldsymbol{X}\boldsymbol{w}]$
  - Are $\boldsymbol{w}^T \boldsymbol{X}^T \boldsymbol{y}$ and $\boldsymbol{y}^T \boldsymbol{X}\boldsymbol{w}$ the same?
  - $= \frac{1}{2} [\boldsymbol{y}^T \boldsymbol{y} - 2\boldsymbol{w}^T \boldsymbol{X}^T \boldsymbol{y} + \boldsymbol{w}^T \boldsymbol{X}^T \boldsymbol{X}\boldsymbol{w}]$

- Compute derivative
  - $\nabla_{\boldsymbol{w}} J(\boldsymbol{w}) = \frac{1}{2} [0 - 2\boldsymbol{X}^T \boldsymbol{y} + 2\boldsymbol{w}^T \boldsymbol{X}^T \boldsymbol{X}]$ Mismatched dimensions!
  - $\nabla_{\boldsymbol{w}} J(\boldsymbol{w}) = \frac{1}{2} [0 - 2\boldsymbol{X}^T \boldsymbol{y} + 2\boldsymbol{X}^T \boldsymbol{X}\boldsymbol{w}]$
  - $= -\boldsymbol{X}^T \boldsymbol{y} + \boldsymbol{X}^T \boldsymbol{X}\boldsymbol{w}$

- Closed form solution:
  - Set equal to zero and solve
  - $-\boldsymbol{X}^T \boldsymbol{y} + \boldsymbol{X}^T \boldsymbol{X}\boldsymbol{w} = 0$
  - $\boldsymbol{X}^T \boldsymbol{X}\boldsymbol{w} = \boldsymbol{X}^T \boldsymbol{y}$ "Normal equation"
  - $\boldsymbol{w} = \left(\boldsymbol{X}^T \boldsymbol{X}\right)^{-1} \boldsymbol{X}^T \boldsymbol{y}$