Monday 9.16.24

Simple Linear Regression: (review)

- Model: $h_{\vec{w}} = W_0 + W_1 x = \hat{y}$
  bias ↑ weight ↑
- Cost Function: $J(W_0, W_1) = \frac{1}{2} \sum_{i=1} (\underbrace{y_i}_{\text{actual } y_i} - \underbrace{W_0 - W_1 x_i}_{\text{Predicted } y_i})^2 = RSS$
  └ measures model performance
  residual
  - $\hat{W}_1 = \dfrac{Cov(x,y)}{Var(x)}$
  - $\hat{W}_0 = \bar{y} - \hat{W}_1 \bar{x}$

Applied Linear Algebra: (intro)

- Vectors:
  - Magnitude: $v = \begin{bmatrix} x \\ y \end{bmatrix}$, then $|v| = \sqrt{x^2+y^2}$
  - Ways to write vector: $v = \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} x & y \end{bmatrix}^T$ ✓ transpose
  - Vector Addition: add corresponding values
    - must have same dimensions
    - $V_1 + V_2 = \begin{bmatrix} 8 \\ 3 \end{bmatrix} + \begin{bmatrix} 1 \\ 5 \end{bmatrix} = \begin{bmatrix} 9 \\ 8 \end{bmatrix}$ (output is new vector)
      └ aka moving V_2 to the end of V_1
      └ New vector: start of V_1 to end of V_2
  - Vector Dot Product: multiply corresponding values, add products
    - $V_1 \cdot V_3 = \begin{bmatrix} 8 \\ 3 \end{bmatrix} \cdot \begin{bmatrix} -3 \\ 8 \end{bmatrix} = -3 \cdot 8 + 8 \cdot 3 = 0$ (output is one number)
      - if dot product is 0, vectors are perpendicular
      - Positive -> "same" direction
- Matrices:
  - Matrix addition: add corresponding values
    - must have same dimensions
    - $A + B = \begin{bmatrix} a & b \\ c & d \end{bmatrix} + \begin{bmatrix} e & f \\ g & h \end{bmatrix} = \begin{bmatrix} a+e & b+f \\ c+g & d+h \end{bmatrix}$ (output is new matrix)
  - Matrix Multiplication: Not Commutative
    - inner dimensions must match
      - A # of Rows = B # of Cols
    - if A shape is (m,n) and B shape is (n,p), AB shape is (m,p)
      - Outer dimensions
    - $AB = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} e & f \\ g & h \end{bmatrix} = \begin{bmatrix} ae+bg & af+bh \\ ce+dg & cf+dh \end{bmatrix}$

- Matrix Transpose:
  - $A = \begin{bmatrix} a & b & c \\ d & e & f \end{bmatrix}$ then $A^T = \begin{bmatrix} a & d \\ b & e \\ c & f \end{bmatrix}$
    - useful note: $(AB)^T = B^T A^T$

- Matrix Inverse:
  - $AA^{-1} = I$ (Identity Matrix)
    - $A\cancel{I} = A$
  - $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ $A^{-1} = \frac{1}{ad-bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$
  - $AA^{-1} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} \frac{d}{ad-bc} & -\frac{b}{ad-bc} \\ \frac{-c}{ad-bc} & \frac{a}{ad-bc} \end{bmatrix} = \begin{bmatrix} \frac{ad - \cancel{bc}}{\cancel{ad} - bc} & \frac{\cancel{-ab + ba}}{ad-bc} \\ \frac{\cancel{cd - dc}}{ad-bc} & \frac{-cb + \cancel{da}}{a\cancel{d} - bc} \end{bmatrix}$

    $= \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = I$

    - can extend infinitely

Multiple Linear Regression: multiple features

$\hat{y} = h_{\vec{w}}(\vec{x}) = W_0 + W_1 X_1 + W_2 X_2 + \ldots + W_p X_p = \vec{w} \cdot \vec{x}$

$x_0 = 1$

want to find

$X = \begin{bmatrix} 1 & X_{11} & X_{12} & X_{13} \ldots & X_{1p} \\ 1 & & & & \\ \vdots & & & & \\ 1 & & & & \end{bmatrix}$

"Fake" ones $n \times (p+1)$

* goal is the same:

$J(\vec{w}) = \frac{1}{2} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$

$= \frac{1}{2} \sum_{i=1}^{n} (y_i - \vec{w} \cdot \vec{x}_i)^2$

- minimize cost function

Computing Predictions given X & W

$X\vec{w} = \begin{bmatrix} 1 & x_{i1} & x_{i2} & \ldots & x_{ip} \\ \vdots & & & & \\ 1 & & & & \end{bmatrix} \begin{bmatrix} w_0 \\ w_1 \\ w_2 \\ \vdots \\ w_{p-1} \\ w_p \end{bmatrix} = \underbrace{[\vec{w} \cdot \vec{x}_i]}_{\hat{y}\text{ (Predictions)}}$

Analytic Solution to multiple Linear regression:

$$J(\vec{w}) = \frac{1}{2}\sum_{i=1}^{n}(y_i - \vec{w}\cdot\vec{x}_i)^2$$

$$= \frac{1}{2}(\vec{y} - X\vec{w})\cdot(\vec{y} - X\vec{w})$$

$$= \frac{1}{2}(\vec{y}\cdot\vec{y} - 2X\vec{w} - X\vec{w}\cdot X\vec{w}$$

$$\frac{\partial J}{\partial \vec{w}} = 0 \quad -X^T\vec{y} + (X^TX)\vec{w} = \vec{0} \quad \begin{bmatrix} 0 \\ 0 \\ \vdots \end{bmatrix}$$

$$(X^TX)\vec{w} = X^T\vec{y}$$

$$\underbrace{(X^TX)^{-1}(X^TX)}_{I}\vec{w} = (X^TX)^{-1}X^T\vec{y}$$

$$\boxed{\vec{w} = (X^TX)^{-1}X^T\vec{y}}$$

$$\vec{w} = Var(X)\,Cov(X, y)$$

$\vec{y}$ shape $n \times 1$

X shape $n \times (p+1)$

$\vec{y}\cdot X$ is a shape mismatch,

↳ thus X^T shape $(p+1) \times n$

Shape: $[(p+1)\times 1] \quad (p+1)\times\cancel{(p+1)}\cdot\cancel{(p+1)}\times 1 = (p+1)\times 1$