Key Machine Learning Concepts

Start Simple First

Don’t Start with Neural Networks!

- Complexity
- Maintenance
- Explainability
- Slow and costly
- Turing’s Rule of Least Power

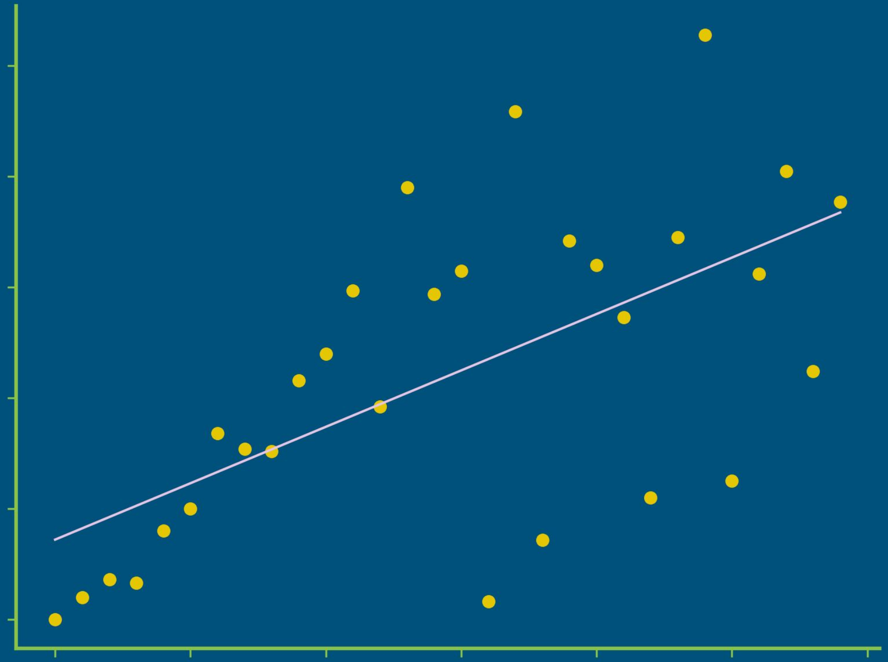
Linear Regression

Model : $y = mx + b$

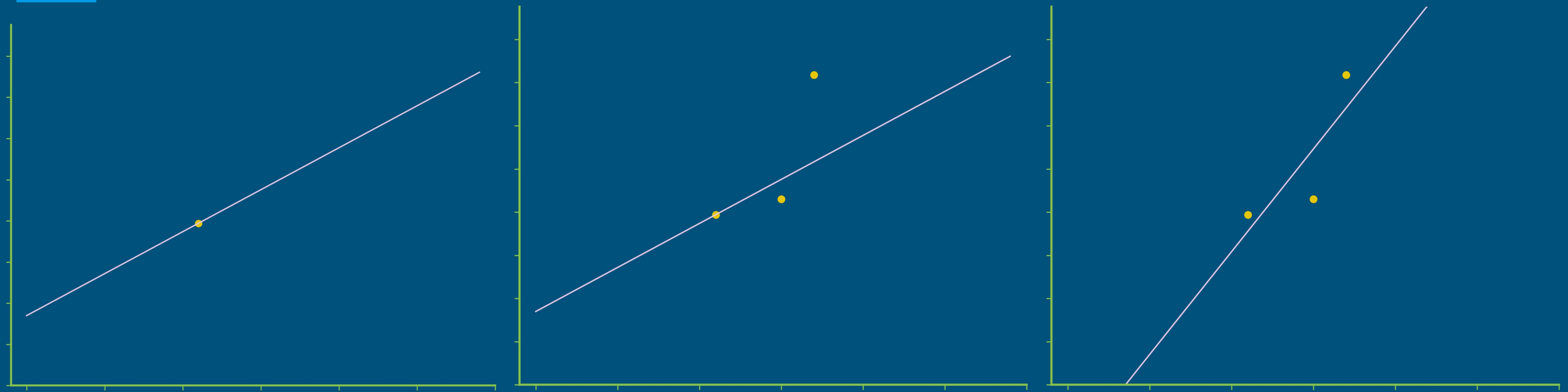
- Supervised
- High Bias
- Low Variance

$m = \sum(x_i - \bar{x})(y_i - \bar{y}) \,/\, \sum(x_i - \bar{x})^2$

$b = \bar{y} - m\bar{x}$



Judging Goodness of Fit



Loss Function : loss = $^{1}/_{2n} \sum ((Model(x) - y)^2)$

$(9^2 + 0^2) / 4 = 81 / 4 = 20.25$

$(4.5^2 + 4.5^2) / 4 = 40.5 / 4 = 10.125$

Gradient Descent

Find partial derivative of each parameter

$m_1 := m_0 - (\textit{Learning Rate} * D_m)$

Optimizer : Minimizes *Loss Function* by updating *Model* parameters

Stochastic Gradient Descent (SGD)

$D_m = {}^{1}/_{2n} \sum ({}^{d}/_{dm} (y - (mx + b))^2)$

$u = y - (mx + b)$

$D_m = {}^{1}/_{2n} \sum ({}^{d}/_{dm} (u^2))$

${}^{d}/_{dm} (u^2) = 2u * {}^{du}/_{dm}$

${}^{du}/_{dm} = {}^{d}/_{dm} (y - (mx + b)) = -x$

${}^{d}/_{dm} = 2(y - (mx + b)) * -x$

$D_m = {}^{1}/_{2n} \sum (2(y - (mx + b))(-x))$

$D_m = {}^{1}/_{n} \sum ((y - (mx + b))(-x))$

Multiple Linear Regression

$y_0 = b + m_0 x_{00} + m_1 x_{01} \dots m_n x_{0n}$

$Y = BX$

$$\begin{bmatrix} y_0 \\ \vdots \\ y_p \end{bmatrix} = \begin{bmatrix} b \\ m_0 \\ \vdots \\ m_n \end{bmatrix} X \begin{bmatrix} 1, & x_{00} \dots & x_{0n} \\ 1, & \vdots & \vdots \\ 1, & x_{p0} \dots & x_{pn} \end{bmatrix}$$

Multiple LR - Gradient Descent

$Loss = {}^{1}/_{2n} (Y - BX)^T(Y - BX)$

$D_B = {}^{1}/_{n} X^T(BX - Y)$

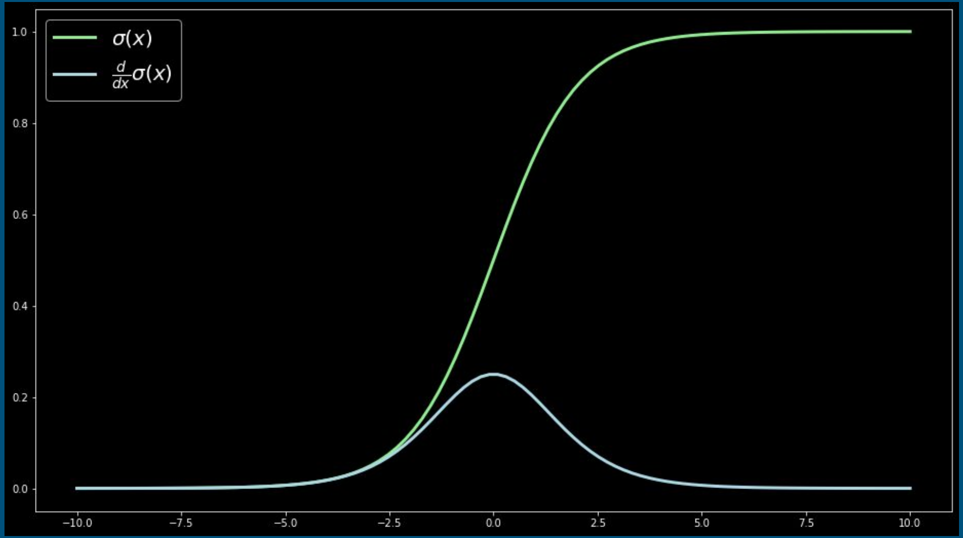
$B1 := B0 - \alpha D_B$

Logistic Regression

Sigmoid *Activation Function* : Condenses output to be between 0 and 1

Sigmoid = $1 / (1 + e^{-x}) = \boldsymbol{\sigma}(x)$

$y = 1 / 1 + e^{-(mx + b)} = \boldsymbol{\sigma}(mx + b)$



Neural Networks

3 layers, 3 neurons per layer, sigmoid activation function

$Y = \boldsymbol{\sigma}(\boldsymbol{\sigma}(\boldsymbol{\sigma}(M_1X + b_1)M_2 + b_2)M_3 + b_3)$

$M_q \in \mathbb{R}^{3 \times n}$

$b_q \in \mathbb{R}^{3 \times 1}$

Principles for Machine Learning

- *Model* : Describes variable relationships : Linear Regression
- *Loss Function* : Describes goodness of fit : Mean Squared Error
- *Optimizer* : Minimizes loss function : Stochastic Gradient Descent (SGD)
- *Hyperparameters* :
    - *Learning Rate* : $\sim 10^{-3} - 10^{-6}$
    - *Batch Size* : n
- *Evaluation Metrics* : Tells human how well model is working : RMSE

Happy Hunting!