Bayes' Theorem

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

<u>Term Explanations</u>

$P(A|B) = \text{posterior probability of } A \text{ given } B$

$P(B|A) = \text{likelihood of } B \text{ given } A$

$P(A) = \text{prior probability of } A$

$P(B) = \text{marginal likelihood of } B$

#1

@aiinminutes

Standardization (Z-score)

$$z_i = \frac{x_i - \mu}{\sigma}$$

Term Explanations

z_i = standardized score for data point x_i

x_i = original data point i

μ = mean of the data

σ = standard deviation of the data

#2

@aiinminutes

Mean Square Error (MSE)

$$\text{MSE} = \frac{1}{m}\sum_{i=1}^{m}(y_i - \hat{y}_i)^2$$

<u>Term Explanations</u>

$\text{MSE} = \text{Mean Square Error}$

$y_i = \text{actual value for data point } i$

$\hat{y}_i = \text{predicted value for data point } i$

$m = \text{number of data points}$

@aiinminutes

#3

Multi-class Cross Entropy Loss

$$L = -\sum_{i=1}^{m}\sum_{k=1}^{K} y_{ik}\log(\hat{p}_{ik})$$

Term Explanations

$L =$ cross entropy loss

$y_{ik} = 1$ if true label of i is k else 0

$\hat{p}_{ik} =$ predicted probability of i being in class k

$m =$ number of data points

$K =$ number of classes

#4

@aiinminutes

Softmax Function

$$\sigma(z_i) = \frac{e^{z_i}}{\sum_{j=1}^{K} e^{z_j}}$$

Term Explanations

$\sigma(z_i)$ = softmax score for class i

e^{z_i} = exponentiated score for class i

$\sum_{j=1}^{K} e^{z_j}$ = sum of exponentiated scores for all classes

#5

@aiinminutes

Gradient Descent Update Rule

$$\mathbf{w} \leftarrow \mathbf{w} - \eta \nabla_{\mathbf{w}} J(\mathbf{w})$$

Term Explanations

$\mathbf{w}$ = weight vector

η = learning rate

$\nabla_{\mathbf{w}} J(\mathbf{w})$ = gradient of the loss function J with respect to $\mathbf{w}$

Linear Regression
Normal Equations (MATRIX)

$$\mathbf{w} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$$

<u>Term Explanations</u>

$\mathbf{w}$ = weight vector

$\mathbf{X}$ = data matrix (features)

$\mathbf{X}^T$ = transpose of data matrix

$\mathbf{y}$ = vector of true values

$(\mathbf{X}^T\mathbf{X})^{-1}$ = inverse of matrix product

@aiinminutes

#7

Logistic Regression

$$\hat{\mathbf{y}} = \sigma(\mathbf{X}\mathbf{w} + \mathbf{b})$$

$$J(\mathbf{w}, \mathbf{b}) = -\frac{1}{m}\left(\mathbf{y}^T \log(\hat{\mathbf{y}}) + (1 - \mathbf{y})^T \log(1 - \hat{\mathbf{y}})\right)$$

Term Explanations

$\hat{\mathbf{y}}$ = vector of predicted probabilities

$\sigma(z)$ = sigmoid function $\sigma(z) = \dfrac{1}{1 + e^{-z}}$

$\mathbf{X}$ = data matrix (features)

$\mathbf{w}$ = weight vector

$\mathbf{b}$ = bias vector

$J(\mathbf{w}, \mathbf{b})$ = cost function

$\mathbf{y}$ = vector of true labels

m = number of data points

@aiinminutes

#8

K-means Clustering Objective

$$J = \sum_{i=1}^{m} \sum_{k=1}^{K} \mathbf{1}_{\{c_i=k\}} \|x_i - \mu_k\|^2$$

Term Explanations

J = total within-cluster sum of squares

$\mathbf{1}_{\{c_i=k\}}$ = indicator function for data point x_i in cluster k

x_i = data point i

μ_k = centroid of cluster k

#9

@aiinminutes

Principal Component Analysis

$$\Sigma = \frac{1}{m} X^T X$$

$$w = \operatorname{argmax}_w \left(w^T \Sigma w \text{ subject to } \|w\| = 1 \right)$$

<u>Term Explanations</u>

$\Sigma =$ covariance matrix

$X =$ data matrix

$m =$ number of data points

$w =$ principal component vector

@aiinminutes

#10