Naïve Bayes Classifier

* Assumption: Features are conditionally independent given the class label
* Probability Calculation: P(Y∣X) ∝ P(X∣Y)P(Y)
* Number of Parameters:
  + With assumption: (k-1)m + (m-1)
  + Without assumption: n(k-1)m + (m-1)

Decision Trees



* Entropy: H(S)= −∑ pi ​log2​(pi​)



* Information Gain (IG): IG(A)=H(S) − ∑(∣Sv∣/∣S∣)​H(Sv​)
  + Where Sv are subsets created by splitting on attribute A



Logistic Regression



* Sigmoid Function: σ(z)=1/(1+e^−z)



* Cross-Entropy Loss: E(w)=−∑y log σ(wTx) + (1−y) log(1−σ(wTx))​
* Gradient Descent Variants:
  + Batch Gradient Descent (BGD): Uses all training data in each step.
  + Stochastic Gradient Descent (SGD): Updates weights for each training sample.
  + Mini-Batch Gradient Descent (MiniBGD): Uses small batches of data.

SoftMax Regression



* SoftMax Function: P(y=i ∣ x)= e^(wT​x)/ ∑j​e^(wT​x)​



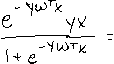
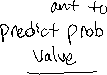
* Gradient Calculation: ∇E(W)=∑ (yi​−P(yi​ ∣ x))x



* SoftMax vs. Sigmoid
  + SoftMax generalizes logistic regression to multiple classes.
  + Logistic regression is a special case of SoftMax for binary classification.

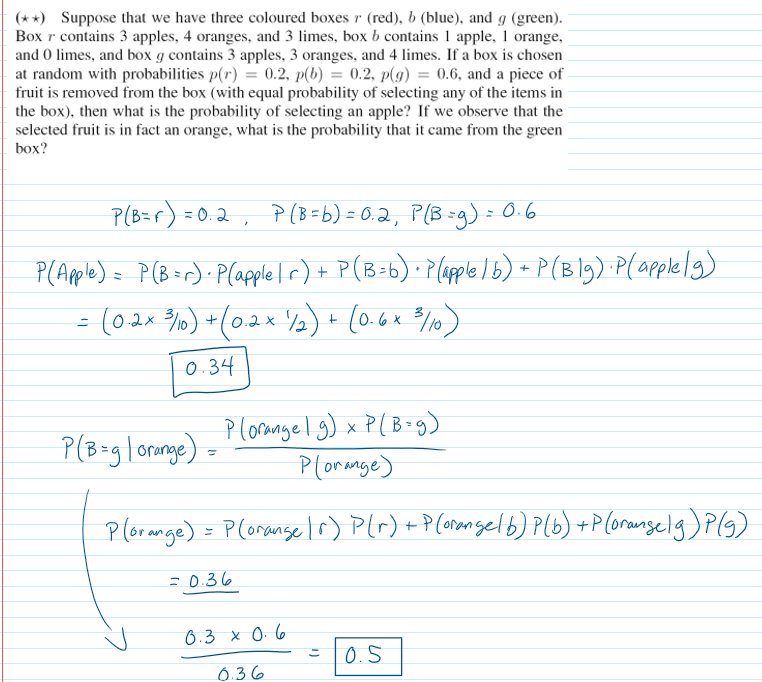
Data Processing

* Splitting Training & Validation Data: Avoids overfitting and helps tune hyperparameters.
* Adding Bias Term: Helps learn an interception in linear models.



The decision boundary is still linear, because logistic regression always results in a linear boundary regardless of the threshold. However, changing the threshold affects the confidence required for classification, making it stricter or looser.

A page of a math test

AI-generated content may be incorrect.A screenshot of a computer

AI-generated content may be incorrect.A paper with writing on it

AI-generated content may be incorrect.

