Bayes Error, Linear and Logistic Regression Classifiers

MACHINE LEARNING UNIT 13

Probabilistic Classifier

- Predicted label : mode of the posterior distribution!
- $Y_{pred} = \text{argmax}_k \, p(Y = k \mid X)$
- Confidence of the prediction = $p(Y = Y_{pred} \mid X)$

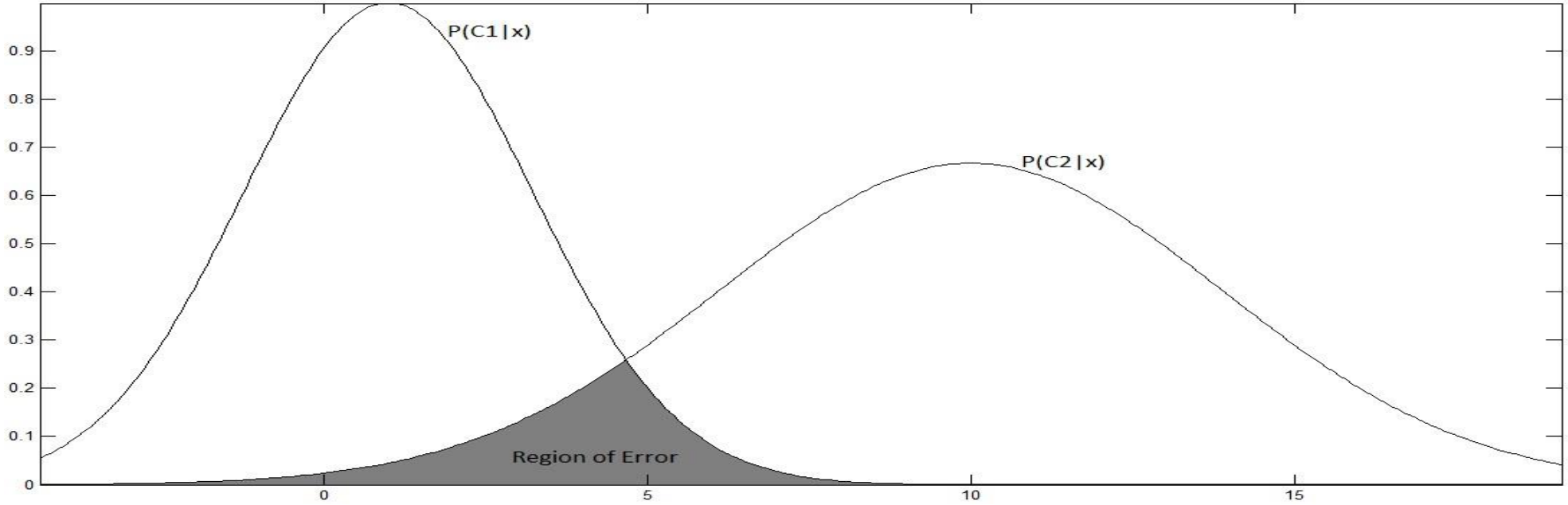
- If Bayesian approach used for $p(Y \mid X)$: Bayesian Classifier!

Error in Bayes Classifier

- Bayes error probability = total probability of the non-mode classes!
- Risk of prediction = 1 – confidence of prediction
- Bayes error = expected risk (expectation over all X and Y)

Error in Bayes Classifier

- Bayes error probability = total probability of the non-mode classes!
- Risk of prediction = 1 – confidence of prediction
- Bayes error = expected risk (expectation over all X and Y)



Bayesian or Frequentist?

- Frequentist approach: Estimate p(Y | X) from data
- Frequentist approach: More robust if some classes are rare
- Frequentist approach: More straightforward

- Bayesian approach: Estimate p(X | Y) and p(Y) from data
- Bayesian approach: More robust if some feature values are rare
- Bayesian approach: More interpretable in many real applications
- Huge advantage for high-dimensional features!!!

Features for classification

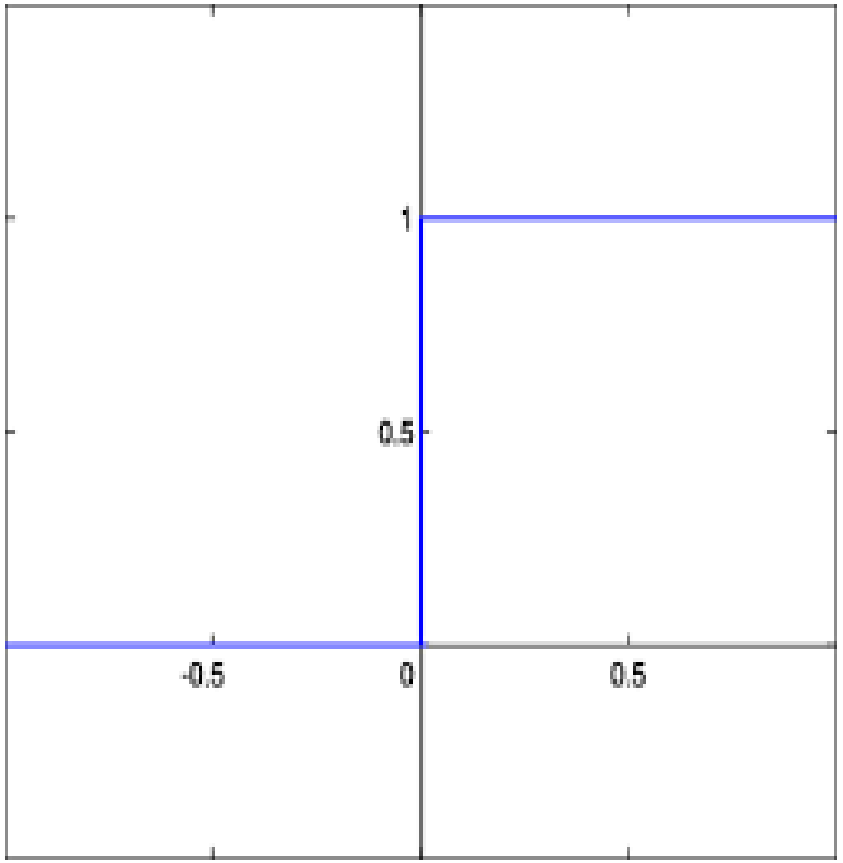
- Non-Ordinal Features (where values have no ordering)
  - Values cannot be plotted on axis
  - eg. Colour of a camera

- Ordinal Features:
  - Values can be plotted on axis
  - eg. Battery life of the camera

- Classification algorithm may depend on nature of features

Threshold-based classification

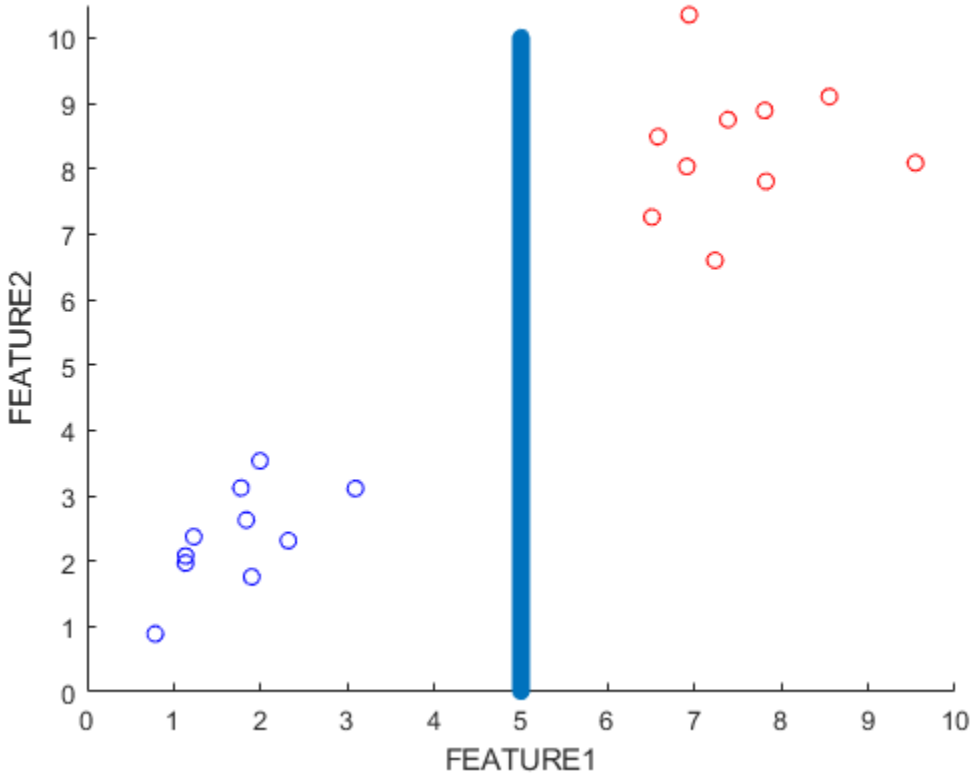
For numeric (integer or real-valued) features, threshold is a simple classifier

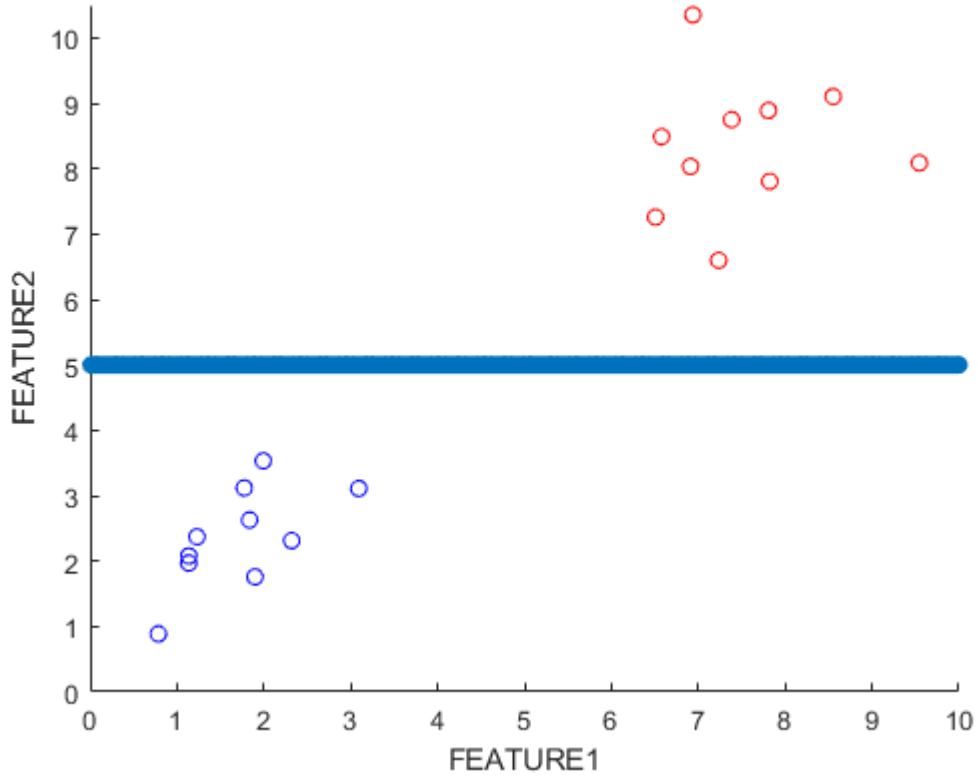


$P(Y = 1 \mid X) = 1$ if $X > 0$

$= 0$ if $X < 0$

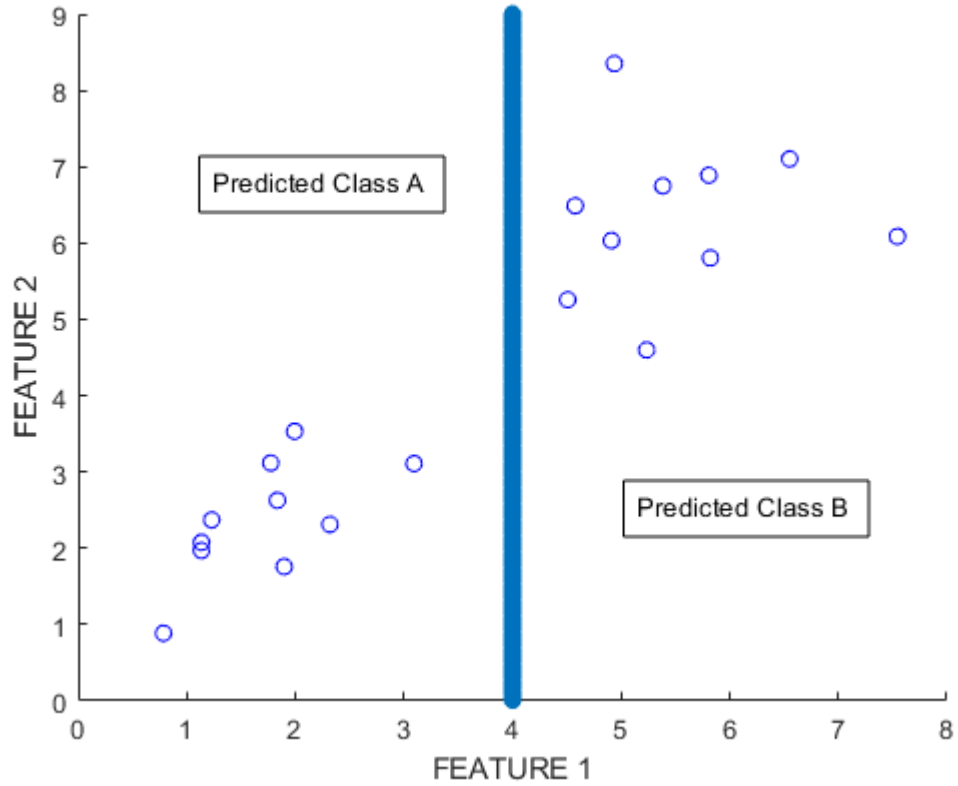
Threshold-based classification





Linear Classifier

- Suitable for ordinal features only
- Linear Classifier: a line/hyperplane in the feature space
- Divides the feature space into 2 parts (half-spaces)
- Data-points in each half-space considered to be of one class



Linear Classifier Notation

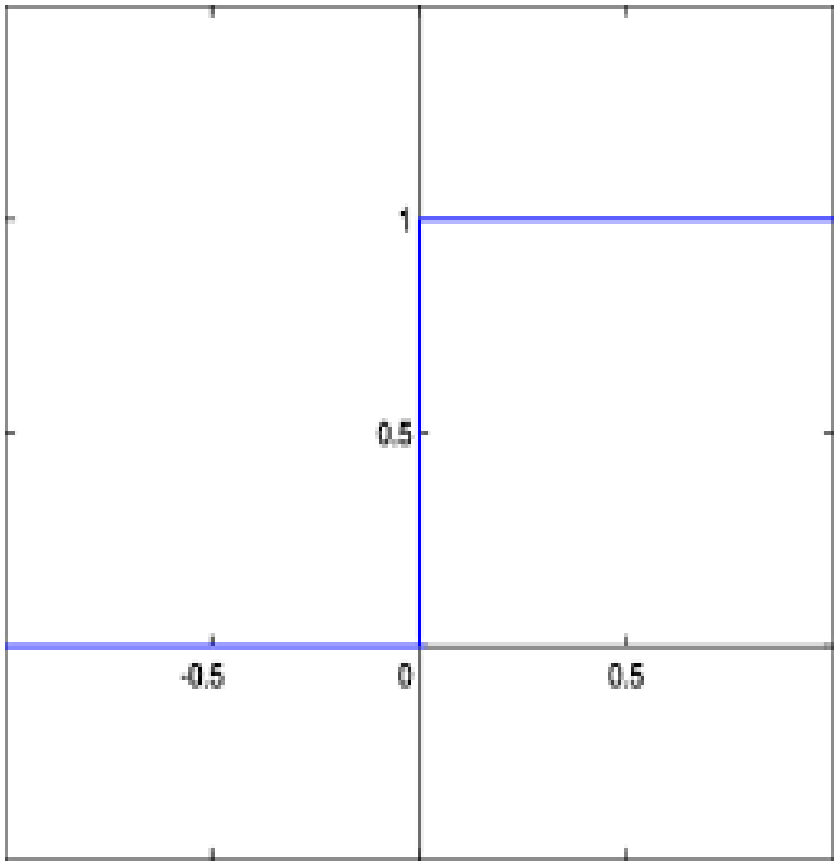
- Hyperplane equation: $W.X + b = 0$
- X: d-dimensional vector in feature space
- W: d-dimensional coefficient vector, b: intercept/bias

Hyperplane equation: $3x_1 + 4x_2 + 5x_3 = 7$

New form: W = [3 4 5], b = -7

- One side of hyperplane: $W.X + b > 0$, other side: $W.X + b < 0$
- Prediction: $y = \text{sign}(W.X + b)$ [2 classes: +1 or -1] (by convention)

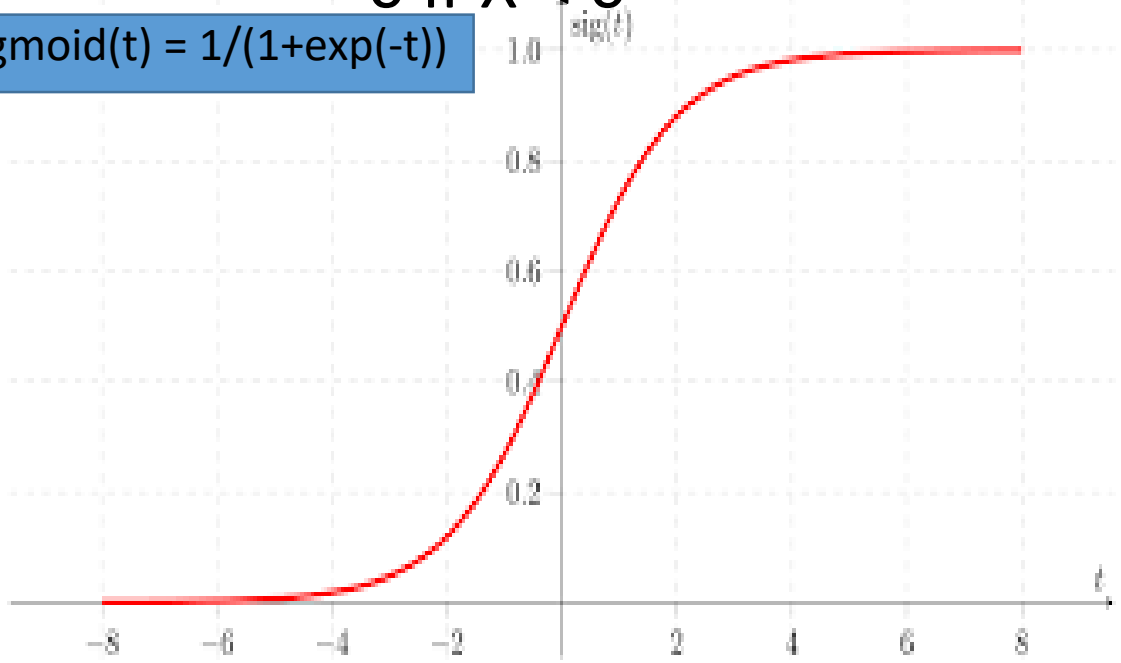
Logistic Regression for Classification



$P(Y = 1 \mid X) = 1$ if $X > 0$

$= 0$ if $X < 0$

sigmoid(t) = 1/(1+exp(-t))



Logistic Regression

- $P(Y = 1 \mid X) = 1$ if $X > 0$

  $= 0$ if $X < 0$
- Approximation: $p(Y = 1 \mid X) = 1/(1 + \exp(-X)) = \sigma(X)$

- Multi-dimensional features: consider weighted combination $w.X$

- $P(Y = 1 \mid X) = 1/(1 + \exp(-w.X)) = \sigma(w.X)$ LOGISTIC

Logistic Regression

- $P(Y = 1 \mid X) = 1$ if $X > 0$
  $= 0$ if $X < 0$
- Approximation: $p(Y = 1 \mid X) = 1/(1 + \exp(-X))$

- Multi-dimensional features: consider weighted combination $w.X$

- $P(Y = 1 \mid X) = 1/(1 + \exp(-w.X))$ LOGISTIC
- But how to find w? REGRESSION!