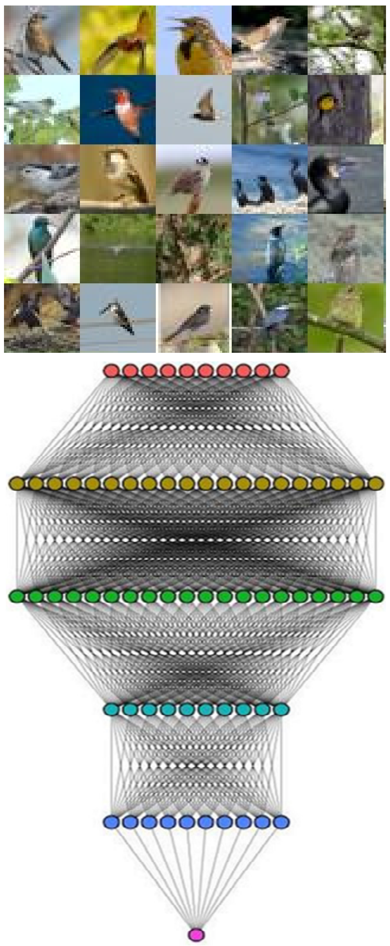
Bayesian Classification

Performance Analysis

Prithwijit Guha
Dept. of EEE, IIT Guwahati

Supervised Learning



- Bayesian Classification, MAP, Chebyshev Inequality
- Performance Measures, Confusion Matrix, ROC Curves
- Logistic Regression
- Perceptron
- Multi-Layer Perceptron (MLP), ELM
- MLP Architectures, Learning, Interpretations
- Non-parametric Methods and K-NN
- Radial Basis Function Neural Networks
- Data Balancing; SMOTE & Weighted Loss Functions
- Classification & Regression Trees
- Support Vector Machines & Multiple Kernel Learning
- Ensemble Methods, Bagging and Boosting

Classification using Chebyshev Inequality

$$S = \{(\boldsymbol{x}_i, y_i); i = 1, \dots n\}$$

$\boldsymbol{x} \in \mathbb{R}^d$

$y \in \{0,1\}$

Classification using Chebyshev Inequality

$$\mu_j = E(\boldsymbol{x}[j]) \qquad v_j = E\left(\boldsymbol{x}[j] - \mu_j\right)^2 \qquad j = 1, \dots d$$

Classification Rule along Dimension k

$$\boldsymbol{R}_k\text{: } [(\boldsymbol{x}[k] - \mu_k)^2 \leq \lambda^2 v_k]$$

$$y(\boldsymbol{x}) = \begin{cases} 0, & \bigwedge_{k=1}^{d} \boldsymbol{R}_k \\ 1, & \text{Otherwise} \end{cases}$$

Bayesian Classification

$$P(\boldsymbol{x}) = \sum_{j=1}^{M} P\left(\boldsymbol{x} \mid \boldsymbol{\omega}_j, \boldsymbol{\theta}_j \right) P(\boldsymbol{\omega}_j)$$

Mixture Model

Bayesian Classification

Likelihood

Prior

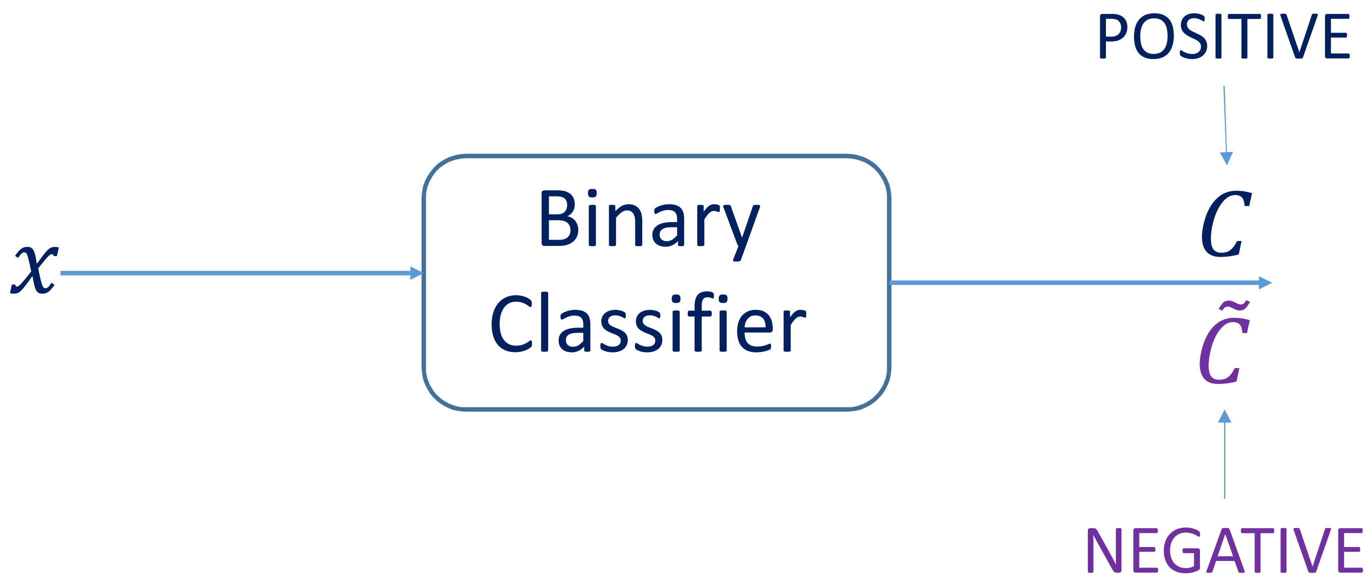
Posterior

$$P\left(\boldsymbol{\omega}_j \mid \boldsymbol{x}\right) = \frac{P\left(\boldsymbol{x} \mid \boldsymbol{\omega}_j, \boldsymbol{\theta}_j\right) P\left(\boldsymbol{\omega}_j\right)}{P(\boldsymbol{x})}$$

Evidence

$$\boxed{P\left(\boldsymbol{\omega}_j \mid \boldsymbol{x}\right) \propto P\left(\boldsymbol{x} \mid \boldsymbol{\omega}_j, \boldsymbol{\theta}_j\right) P\left(\boldsymbol{\omega}_j\right)}$$

Binary Classification Problem



Binary Confusion Matrix

Classification Result

Ground Truth

	C	$\tilde{C}$
C	TP	FN
$\tilde{C}$	FP	TN

Binary Confusion Matrix: Example

Classification Result

Ground Truth

	C	$\tilde{C}$	Total
C	80	20	100
$\tilde{C}$	30	70	100

Performance Measures (TPR)

Sensitivity, Recall, Hit Rate, True Positive Rate (TPR)

$$TPR = \frac{TP}{TP + FN}$$

Example

$$TPR = \frac{80}{80 + 20} = 0.8$$

Performance Measures (TNR)

Specificity, Selectivity, True Negative Rate (TNR)

$$TNR = \frac{TN}{TN + FP}$$

Example

$$TNR = \frac{70}{70 + 30} = 0.7$$

Performance Measures (PPV)

Precision, Positive Predictive Value (PPV)

$$PPV = \frac{TP}{TP + FP}$$

Example

$$PPV = \frac{80}{80 + 30} = 0.727$$

Performance Measures (NPV)

Negative Predictive Value (NPV)

$$NPV = \frac{TN}{TN + FN}$$

Example

$$NPV = \frac{70}{70 + 20} = 0.778$$

Performance Measures (FPR)

Fall Out, False Positive Rate (FPR)

$$FPR = \frac{FP}{TN + FP}$$

Example

$$FPR = \frac{30}{70 + 30} = 0.3$$

Performance Measures (ACC)

Accuracy (ACC)

$$ACC = \frac{TP + TN}{TP + FN + TN + FP}$$

Example

$$ACC = \frac{80 + 70}{100 + 100} = 0.75$$

Performance Measures (BA)

Balanced Accuracy (BA)

$$BA = \frac{TPR + TNR}{2}$$

Example

$$BA = \frac{0.8 + 0.7}{2} = 0.75$$

Performance Measures (F1)

F1 Score (F1)

$$F1^{(p)} = \frac{2 \times PPV \times TPR}{PPV + TPR}$$

$$F1^{(n)} = \frac{2 \times NPV \times TNR}{NPV + TNR}$$

Example

$$F1^{(p)} = \frac{2 \times 0.727 \times 0.8}{0.727 + 0.8} = 0.762$$

$$F1^{(n)} = \frac{2 \times 0.778 \times 0.7}{0.778 + 0.7} = 0.737$$

ROC Curves

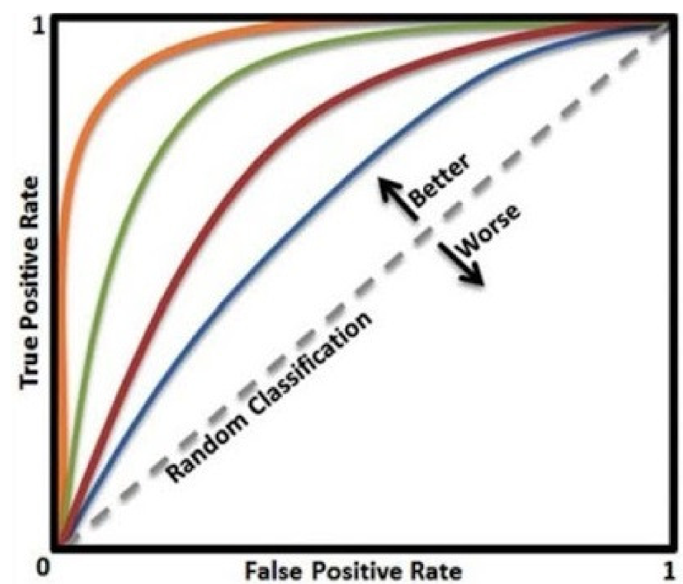
Receiver Operating Characteristics Curves

Example $S = \{(\boldsymbol{x}_i, y_i); i = 1, \dots n\}$

Vary λ

Chebyshev Inequality based Classification

Varying TPR & FPR



Multi-Category Classification Problems

M Category Classification Problem $\quad C_1, C_2, \ldots C_j, \ldots C_M$

CM : Confusion Matrix

$CM[i][j]$: Entity of C_i Detected as that of C_j

Example

C_1	C_2	C_3	C_4	C_5
100	100	100	100	100

Confusion Matrix

Classification Result

Ground Truth

	C_1	C_2	C_3	C_4	C_5	**Total**
C_1	60	25	0	5	10	100
C_2	10	50	10	20	10	100
C_3	5	5	80	0	10	100
C_4	40	20	10	30	0	100
C_5	20	50	10	0	20	100

Performance Measures (ACC)

Overall Accuracy (ACC)

$$ACC = \frac{\sum_{j=1}^{M} CM[j][j]}{\sum_{i=1}^{M} \sum_{j=1}^{M} CM[i][j]}$$

Example

$$ACC = \frac{60 + 50 + 80 + 30 + 20}{100 + 100 + 100 + 100 + 100} = 0.48$$

Reporting Performances

$$\overbrace{\boxed{D_1}\ \boxed{D_2}\ \boxed{D_3}\ \boxed{D_4}\ \boxed{D_5}}^{\text{Total Training Data Set}}$$

$D_1, D_2, D_3, D_4, D_5 \quad A_1$

$D_2, D_3, D_4, D_5, D_1 \quad A_2$

$D_3, D_4, D_5, D_1, D_2 \quad A_3$

$D_4, D_5, D_1, D_2, D_3 \quad A_4$

$D_5, D_1, D_2, D_3, D_4 \quad A_5$

Summary

- Chebyshev Inequality as Classifier
- Bayesian Classification
- The MAP Rule
- Confusion Matrix & Performance Measures
- The ROC Curve

Thank You