Linear Discriminant Analysis (LDA)

Wednesday, September 8, 2021 11:10 AM

- Motivation
  - Logistic Regression → Limited to Binary Resp
    - Created a linear model of : Conditional Dist $P(Y|X)$
  - LDA → Model $K>2$ multiple classes
    - Model $P(X|Y)$ and flip using Bayes Theorem!

* Bayes Rule:

$$P(\text{Cause} \mid \text{Effect}) = \frac{P(\text{Effect} \mid \text{Cause})\, P(\text{Cause})}{P(\text{Effect})}$$

$$P(C,E) = P(C|E)\,P(E)$$

Joint Dist

Conditional Dist

Marginal Dist

$$P(E,C) = P(E|C)\,P(C)$$

$$P(C,E) = P(E,C)$$

- Given:

  K classes in our data: $\{ c_1 , \ldots c_K \}$

X continuous predictors: $X_1 \dots X_p$

π_k overall prior probability that $Y = c_k$

$\leftarrow \frac{\# c_k}{\# \text{Total}}$

pdf/pmf

$$f_k(x) = P(X = x \mid Y = c_k)$$

Likelihood

$$P(Y = c_k \mid X = x) = \frac{f_k(x)\,\pi_k}{\sum_{i=1}^{K} \pi_i f_i(x)}$$

Prior

Posterior

* Prediction

$$\begin{cases} P(Y_0 = c_1 \mid X_0) \\ P(Y_0 = c_2 \mid X_0) \\ \vdots \\ P(Y_0 = c_K \mid X_0) \end{cases}$$

$\leftarrow$ Pick $Y = c_i$ with highest prob

MAP Estimate (Mode of Posterior Dist)

\# Assumptions for LDA:

- $f_k(x)$ is Gaussian!

$$f_k(x) = \frac{1}{\sqrt{2\pi}\,\sigma_k} e^{-\frac{(x-\mu_k)^2}{2\sigma_k^2}}$$

μ_k, σ_k^2 are mean/variance of x for class ck

- We assume: $\sigma_1^2 = \sigma_2^2 = \dots = \sigma_k^2$

$$\therefore \; p_k(x) = \frac{\left(\frac{1}{\sqrt{2\pi}\sigma_k} e^{\frac{-(x-\mu_k)^2}{2\sigma_k^2}}\right)(\pi_k)}{\left(\sum_{i=1}^{K} \pi_i \frac{1}{\sqrt{2\pi}\sigma_k} e^{\frac{-(x-\mu_k)^2}{2\sigma_k^2}}\right)}$$

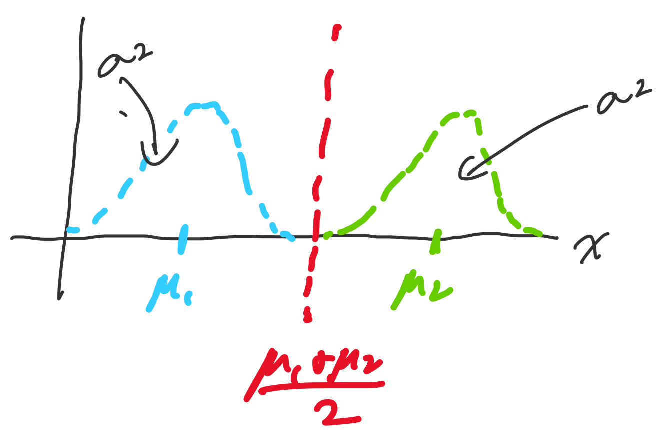
- Discriminant Function:
  - Linear function

$$\delta_k(x) = x\frac{\mu_k}{\sigma^2} - \frac{\mu_k^2}{2\sigma^2} + \log(\pi_k)$$

calculate for each $c_1, c_2 \dots c_k$, pick largest value of $\delta_k(x)$

Ex. $K=2$, $\pi_1 = \pi_2$

$$x = \frac{\mu_1^2 - \mu_2^2}{2(\mu_1 - \mu_2)} = \frac{\mu_1 + \mu_2}{2}$$



* LDA with Multivariate Gaussian → $x_1 \cdots x_p$

$$f(x) = \frac{1}{(2\pi)^{p/2} |\Sigma|^{1/2}} e^{-\frac{1}{2}(x-\mu)^T \Sigma^{-1} (x-\mu)}$$

Σ → $\mathrm{cov}(x)$ common to all classes!

→ QDA diagonalizes Σ

- Generative vs. Discriminative Classifiers
  - Logistic Regression → Directly estimate $P(Y|X)$

    Discriminative ↳ $\log \frac{P(x)}{1-P(x)}$

  - LDA → Directly estimating $P(X|Y)$, $P(Y)$

    Generative ↑ assumption of Gaussian form

    Along with Naive Bayes!

    $P(X|Y)P(Y) \rightarrow P(X,Y)$

    from joint $P(X,Y) \rightarrow P(Y|X)$

- Evaluation
  - Training vs Test error rates

    ↳ $E[e_{oos}] = ?$

    ↑ Training error e_{train}

    * overfitting!

$E[e_{oos}] = \,:$

← Point estimate from test set

e_{Test} from test set

$E[e_{oos}] = \,?$

← Dist estimate from cross-validation

e_{CV} from k CV sets

- Confusion Matrix

		Actual	
		+	−
Predicted	+	TP	FP Type I
	−	FN Type II	TN

* COST!

- Metrics

Accuracy: $\frac{TP + TN}{Total}$ $Total = TP + TN + FP + FN$

Error Rate: $\frac{FP + FN}{Total} = 1 - \text{Accuracy}$

- Rates

$TPR = \frac{TP}{TP + FN} = 1 - FNR$ Sensitivity (Recall)

$FNR = \frac{FN}{FN + TP} = 1 - TPR$ Miss Rate

$TNR = \frac{TN}{\;} = 1 - FPR$ Specificity

$TNR = \frac{TN}{TN+FP} = 1 - FPR$ Specificity

$FPR = \frac{FP}{FP+TN} = 1 - TNR$ 1- Specificity (Fall-out)

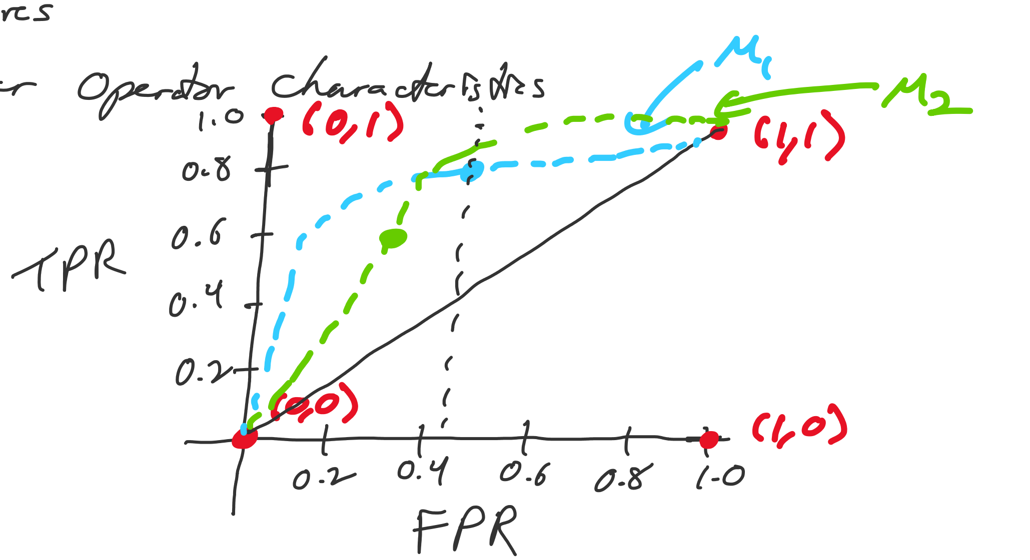
- Values

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

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

* $F_1 = \frac{2}{\frac{1}{p} + \frac{1}{r}}$ ← single number (harmonic mean)

- ROC Curves
  - Receiver Operator Characteristics



* AUC > 0.5