

Aprendizado de Máquina 2019.1

Avaliação de Classificadores

Thiago Ferreira Covões

(slides baseados no material do Prof. Carlos Santos e Prof. Eduardo Hruschka [erh@icmc.usp.br])

Mensurando acerto/erro

- É intuitivo considerar a taxa de acerto de classificação (ou erro correspondente)

$$\frac{TN + TP}{N}$$

- Matriz de confusão (2 classes)

Classe	Predita positiva	Predita negativa
Real Positiva	Verdadeiro Positivo (TP)	Falso Negativo (FN)
Real Negativa	Falso Positivo (FP)	Verdadeiro Negativo (TN)

Mensurando o acerto/erro

- Taxa de acerto pode não ser o ideal
- Os dois classificadores abaixo têm a mesma taxa de acerto. No entanto, um deles parece melhor, qual?

Classe	PPos	PNeg
Pos	0	10
Neg	0	9.990

Classe	PPos	PPeg
Pos	10	0
Neg	10	9.980

Classe menos frequente

- A classe menos frequente (normalmente chamada de positiva) é geralmente a de maior interesse:
 - Doença rara
 - Transação fraudulenta
- Um classificador que erra muito a classe positiva é de pouca utilidade

Custos

- Uma outra maneira é atribuir um custo diferente para cada tipo de erro
- Para calcular o custo total, os erros são multiplicados pelo seu custo
- Em vez de minimizar o erro, o objetivo do classificador é minimizar o custo total.

Custos

- Nem todo classificador consegue incorporar custos na indução de modelos
- É muito difícil fazer uma atribuição de custos
- Custos podem variar com o tempo
- Existe uma relação direta entre custos e alterar artificialmente a proporção de exemplos entre as classes

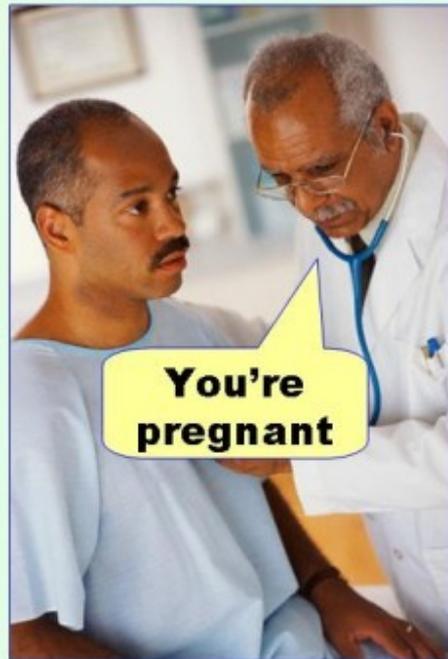
Custos

- É possível incorporar custos:
 - Pesos diferentes para instâncias
 - Reamostragem do conjunto de treinamento, proporção de exemplos positivos/negativos de acordo com custo.

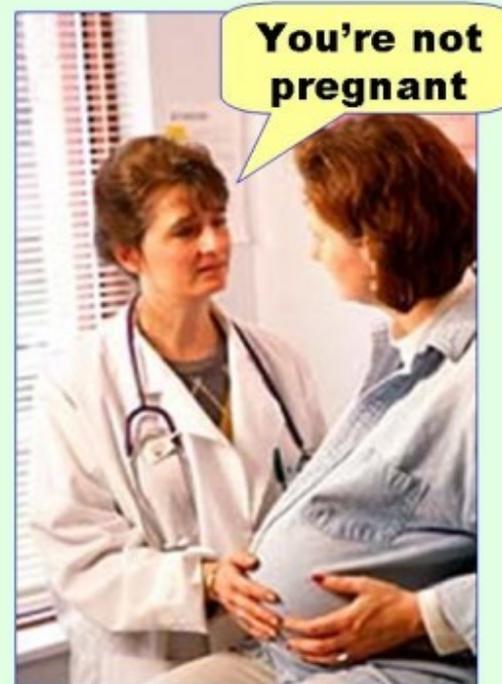
Mensurando erro

- Taxa de Falsos Positivos (Erro Tipo I):
 - De todos os exemplos negativos, quantos foram erroneamente preditos como positivos?
 - $FP/(FP+TN)$
- Taxa de Falsos Negativos (Erro Tipo II):
 - De todos os exemplos positivos, quantos foram erroneamente preditos como negativos?
 - $FN/(FN+TP)$

Type I error
(false positive)



Type II error
(false negative)



Mensurando o erro

- Taxa de Verdadeiros Positivos
 - De todos os objetos positivos, quantos eu acertei?
 - $VP/(VP+FN)$
 - Também conhecido como recall, revocação, sensibilidade

Mensurando o erro

- Precisão
 - Dos objetos preditos como positivos, quantos eram de fato positivos
 - $TP/(TP+FP)$
- F-Measure
 - Media harmônica entre Precisão e Revocação
 - $2 \cdot \frac{\text{precisão} \cdot \text{revocação}}{\text{precisão} + \text{revocação}}$

Exemplo

- Computar taxa de acerto, precisão, revocação, F-measure, taxa de falso positivo e taxa de falso negativo da matriz de confusão abaixo

Classe	Predita positiva	Predita negativa
Real Positiva	70	30
Real Negativa	40	60

Exemplo

- Taxa de acerto= $(TP+TN)/N=$
- Precisão= $TP/(TP+FP)=$
- Revocação= $TP/(TP+FN)=$
- TFP= $FP/(FP+TN)=$
- TFN= $FN/(FN+TP)=$

Classe	Predita positiva	Predita negativa
Real Positiva	70	30
Real Negativa	40	60

Exemplo

- Taxa de acerto= $(TP+TN)/N=130/200=0,65$
- Precisão= $TP/(TP+FP)=70/110=0,63$
- Revocação= $TP/(TP+FN)=70/100=0,7$
- TFP= $FP/(FP+TN)=40/100=0,4$
- TFN= $FN/(FN+TP)=30/100=0,3$

Classe	Predita positiva	Predita negativa
Real Positiva	70	30
Real Negativa	40	60

Classificação versus Ranking

- Alguns classificadores indicam apenas a classe predita
- Outros fornecem um indicativo de confiança (chamados de *score*)
 - Pode ser uma probabilidade
 - Normalmente obtidos/transformados entre $[0,1]$ ou $[0,100]$

Classificação versus Ranking

- Como obter *score* para:
 - *KNN*
 - Árvore de Decisão

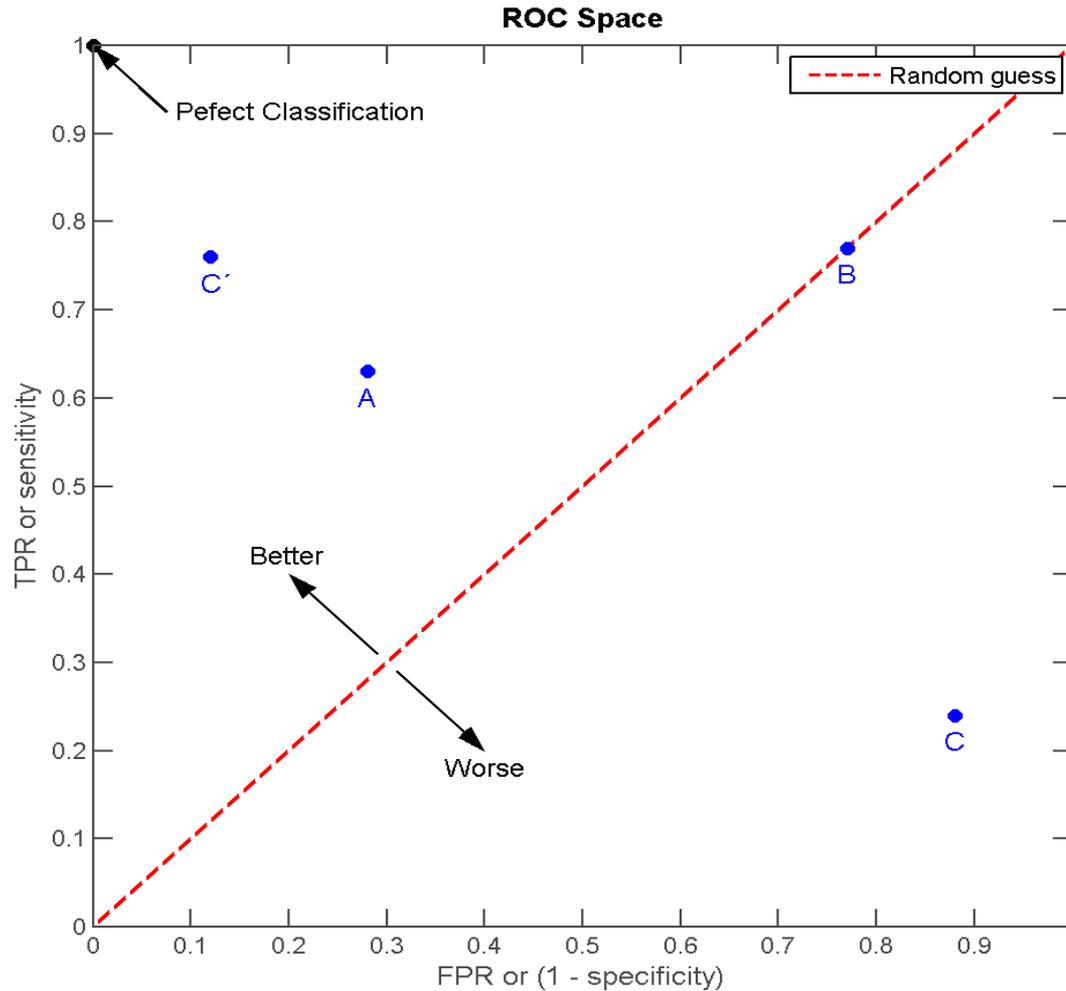
Classificação versus Ranking

- A partir destes *scores* podemos obter diferentes classificadores considerando os valores possíveis para o limiar de decisão
 - Conservador → limiar alto
 - Liberal → limiar baixo

Curvas ROC

- Receiver Operating Characteristic curve
- Origem:
 - detecção de sinais
 - compromisso entre alarme falso/acerto
- TFP x TVP [FPR x TPR]
- $TFP = FP / (FP + TN)$
- $TVP = VP / (VP + FN)$

Curvas ROC



fonte: http://upload.wikimedia.org/wikipedia/commons/3/36/ROC_space-2.png

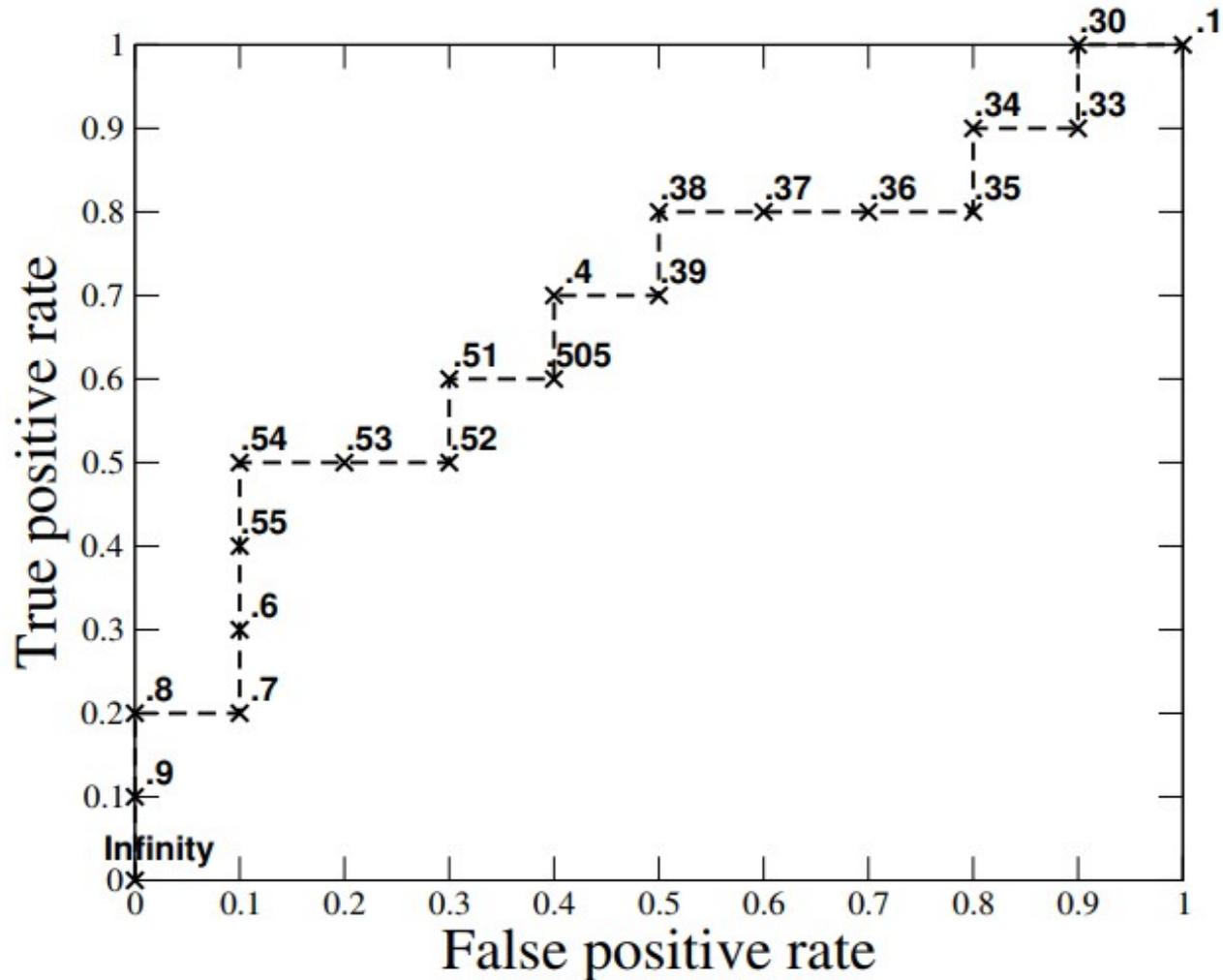
Gerando a curva ROC

- (1) Ordene as tuplas da base de testes por ordem crescente de seus valores de output (prob. de estar na classe positiva)
- (2) Selecione a primeira tupla X1 e
 - (i) Classifique X1 como **POSITIVA**
 - (ii) Classifique todas as tuplas com outputs maiores do X1 como **POSITIVAS**
Neste caso, todas as tuplas foram classificadas como positivas.
Logo: todas as positivas corretamente classificadas $TPR = 1$
todas as negativas incorretamente classificadas $FPR = 1$
- (3) Selecione a segunda tupla X2
 - (i) Classifique X2 como **POSITIVA**
 - (ii) Classifique todas as tuplas com outputs maiores do X2 como **POSITIVAS** e as com outputs menores como **NEGATIVAS**
 - (iii) **Calcule os novos valores de TP e FP**
 - (1) Se a classe de X1 é positiva então TP é decrementado de 1 e FP continua o mesmo
 - (2) Se a classe de X1 é negativa então TP continua o mesmo e FP é decrementado.
- (4) Repita o processo para a terceira tupla até varrer todo o banco de testes
- (5) Faça o gráfico dos valores de TPR (eixo y) por FPR (eixo x)

Gerando a curva ROC

Inst#	Class	Score	Inst#	Class	Score
1	p	.9	11	p	.4
2	p	.8	12	n	.39
3	n	.7	13	p	.38
4	p	.6	14	n	.37
5	p	.55	15	n	.36
6	p	.54	16	n	.35
7	n	.53	17	p	.34
8	n	.52	18	n	.33
9	p	.51	19	p	.30
10	n	.505	20	n	.1

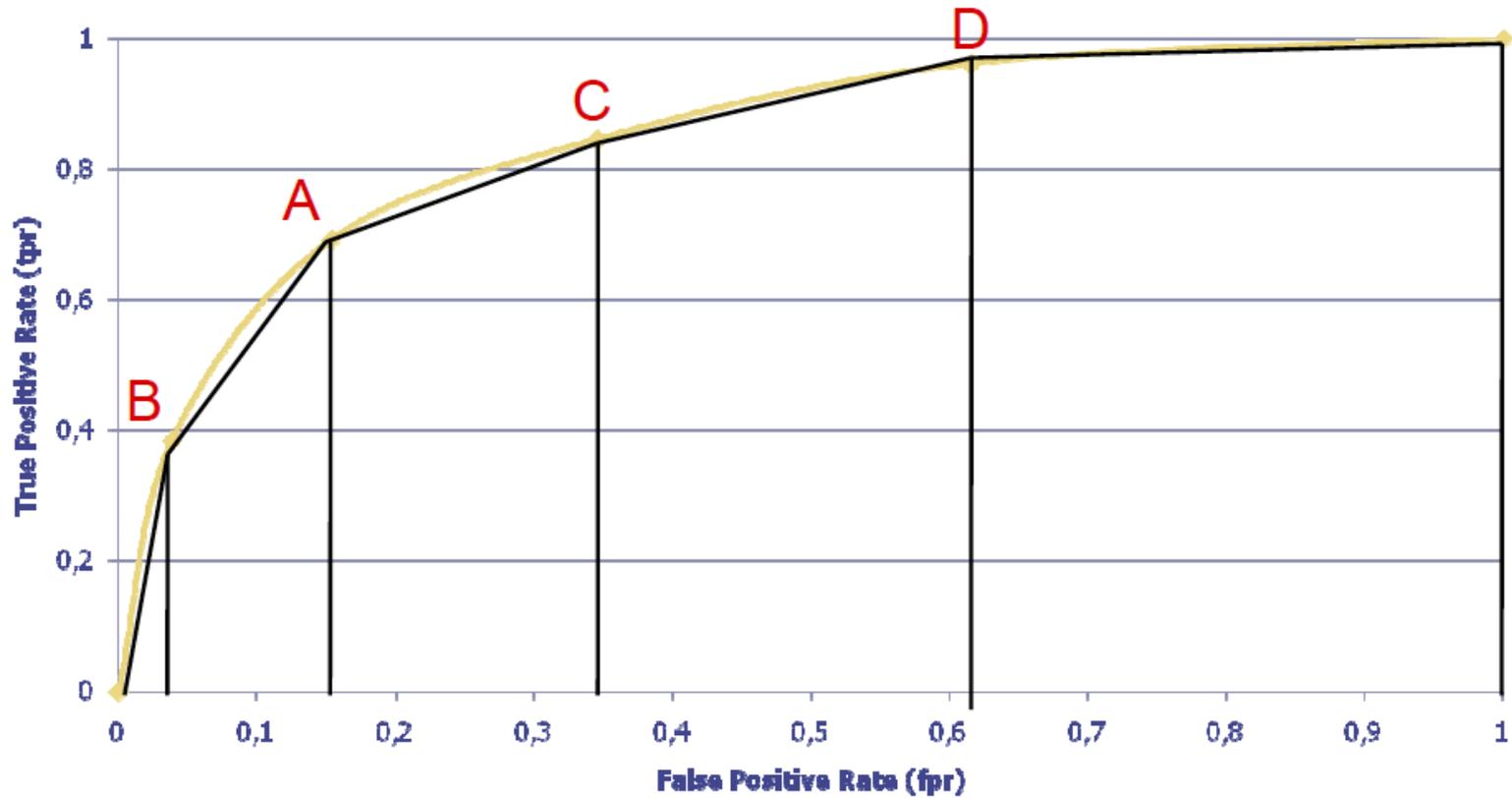
Gerando a curva ROC



AUC

- Probabilidade que um exemplo positivo vai estar ranqueado acima de um exemplo negativo.
- Pode ser calculado pela regra do trapézio.
- Quanto maior a área, melhor é o desempenho médio do classificador.

AUC



Outras formas de avaliação

- Curvas precisão x revocação
- Curvas de custo
- Curvas lift
- ...

Referências

- E. Alpaydin, Introduction to Machine Learning.
- Janez Demšar. 2006. Statistical Comparisons of Classifiers over Multiple Data Sets. J. Mach. Learn. Res. 7 (December 2006), 1-30.