

Classificação

Advanced Institute for Artificial Intelligence

<https://advancedinstitute.ai>

Agenda

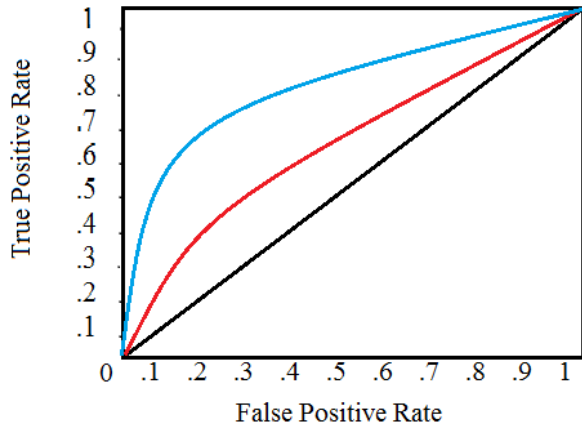
- Curva ROC
- Regressão Logística
- Naive Bayes
- Paradigma de classificação
- Transformando classificador binário em classificador multiclasse
- KNN
- Arvore de Decisão

Curva ROC (*Receiver Operator Characteristic*)

- A relação entre sensibilidade e especificidade pode ser ilustrada usando-se um gráfico conhecido como curva ROC.
- Uma curva ROC é um gráfico de linha que mostra a probabilidade de um resultado positivo verdadeiro comparado a probabilidade de um resultado falso positivo para uma série.

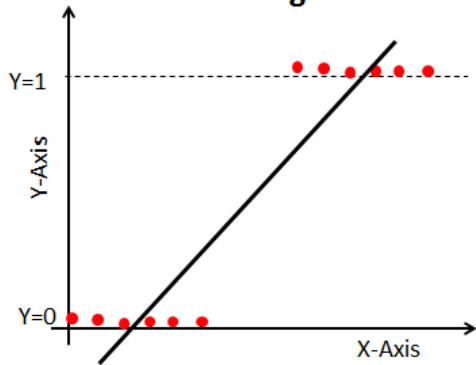
Classificação

Scikit prove um método para curva Roc from `sklearn.metrics import roc_auc_score`

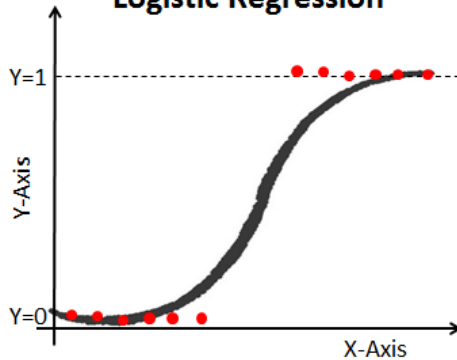


- A regressão logística binária é um tipo de análise de regressão em que a variável dependente é uma variável qualitativa: 0 ou 1
- Uma função gerada por uma regressão linear considerando uma variável qualitativa 0 ou 1, gera valores fora do intervalo 0 ou 1

Linear Regression



Logistic Regression



Regressão logística

- Modelar a probabilidade de um evento ocorrer dependendo dos valores das variáveis independentes.
- Estimar a probabilidade de um evento ocorrer (e também de não ocorrer) para uma dada observação
- Distribuição discreta de espaço amostral 0,1 que tem probabilidade de sucesso p e falha $q = 1 - p$

Regressão logística

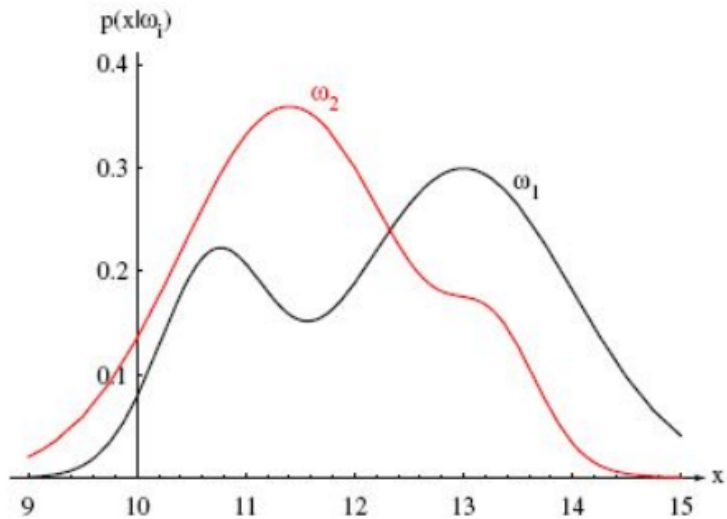
- Na regressão logística estimamos p para qualquer combinação linear das variáveis independentes.
- Isso pode ser alcançado usando o MLE (*Maximum Likelihood Estimation*) para estimar os coeficiente do modelo.
- Também pode ser usado descida do gradiente, otimizando os coeficientes para aproximar os valores mais próximos de 0 e 1.

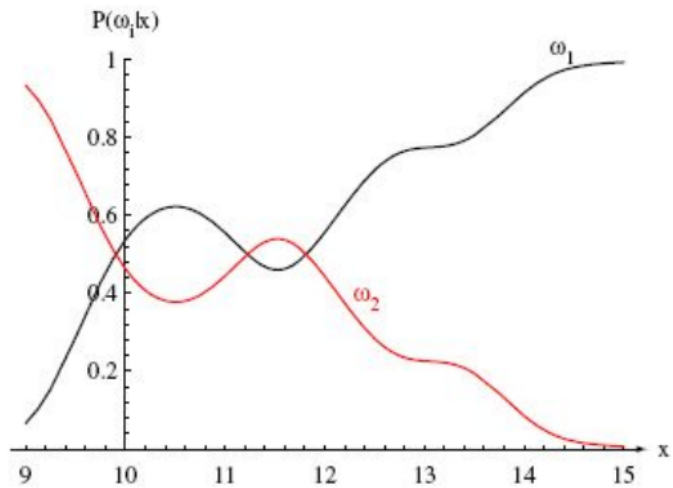
Naive Bayes

- Baseado na suposição de que as quantidades de interesse são reguladas por distribuições de probabilidades.
- Quantificar o custo/benefício entre diferentes decisões de classificação usando probabilidades e custos associados à classificação.
- Teorema de Bayes Mostra como alterar as probabilidades *a priori* tendo em conta novas evidências de forma a obter probabilidades *a posteriori*

- Classe W_1
- Probabilidades a priori $P(W_1)$: Conhecimento a priori que se tem sobre o problema, ou seja, conhecimento a priori sobre a aparição de exemplos das classes do problema.
- Função de Densidade Probabilidade $P(x)$: Frequência com a qual encontramos uma determinada característica (Evidências)

- Densidade de Probabilidade Condicional
- $P(X|W_j)$ (Likelihood) - Verossimilhança
- Frequência com que encontramos uma determinada característica x dado que a mesma pertence a classe W_j
- Densidade de duas classes em que x representa uma característica qualquer





- Probabilidades a posteriori para um valor de $x = 14$,
- a probabilidade do padrão pertencer a W_1 é de 0,08,
- a probabilidade do padrão pertencer a W_2 é de 0,92.
- Para cada x , as probabilidades a posteriori somam 1.

- Um dos algoritmos de aprendizagem mais práticos e utilizados na literatura.
- Denominado Naive (ingênuo) por assumir que os atributos são condicionalmente independentes, ou seja, a informação de um evento não é informativa sobre nenhum outro.
- Apesar dessa premissa, o classificador reporta bom desempenho em diversas tarefas de classificação onde há dependência.

- Aplica-se a tarefas de aprendizagem onde cada instância x é descrita por uma conjunção de valores de atributos em que a função alvo, $F(x)$, pode assumir qualquer valor de um conjunto V
- Um conjunto de exemplos de treinamento da função alvo é fornecido. E então uma nova instância é apresentada, descrita pela tupla de valores de atributos a_1, a_2, \dots
- A tarefa é prever o valor alvo (ou classificação) para esta nova instância.
- Para atributos contínuos o classificador assume que a distribuição de probabilidades dos atributos é normal

Paradigmas de classificação

- Binário (*Binary*): resposta é uma das duas classes possíveis
- Múltiplas Classes (*Multi-Class*): resposta é uma das múltiplas classes possíveis
- Múltiplos Rótulos (*Multi-Label*): resposta é uma ou mais classes possíveis

Utilizando classificador binário para classificação multiclasse:

- Classificação um contra todos (*one versus all*)
- Classificação por todos os pares (*one versus one*)

Classificação um contra todos:

- Cada classificador prevê se a instância pertence à classe de destino ou não
- A classe a qual a instancia pertence retornara como verdadeira, enquanto as outras retornaram como falso
- E se mais de uma classe retornar como verdadeira?
- E se nenhuma classe retornar como verdadeira?

Classificação um contra todos:

- No lugar de usar apenas a previsão binária final de cada classificador, considere o *score* associado a cada previsão.
- Cada classificador geralmente possui alguma medida de *score*
- Utilize a decisão do classificador que apresenta *score* mais alto (maior confiança na previsão)

Classificação por todos os pares:

- Essa abordagem treina um classificador binário para cada par de classes
- A classe que ganhar mais aos pares será a previsão final

- Classificação por todos os pares são mais rápidos para treinar
- Um contra todos é mais rápido em fazer previsões
- O scikit-learn implementa um contra todos por padrão quando você dá mais de duas classes para um classificador binário

- A regressão logística para múltiplos rótulos é chamada regressão logística Multinomial (ou multivariada)
- Baseado em uma função mais geral (softmax) para a probabilidade de múltiplas classes

Regressão logística binomial

- Um Vetor de peso w
- Pontuação conectada à função logística para obter valor entre $[0, 1]$
- Probabilidade de classe negativa é 1 menos probabilidade de classe positiva

Regressão logística Multinomial

- Múltiplos Vetores de peso w_k
- Os scores de vetor de K conectados à função softmax obtêm o vetor de valores de K, cada um entre $[0,1]$ e todos os valores somam 1
- Probabilidade de cada classe depende do seu próprio score do seu vetor de peso

Classificador de árvore de decisão

- Abordagem sistemática para a classificação em várias classes.
- Associa um conjunto de perguntas para o conjunto de dados (relacionado a seus atributos / características)
- Na raiz e em cada um dos nós internos, uma pergunta é feita e os dados nesse nó são divididos em registros separados com características diferentes.
- As folhas da árvore referem-se às classes nas quais o conjunto de dados está dividido.

Classificador KNN (k-vizinhos mais próximos)

- O algoritmo de classificação KNN independe da estrutura dos dados.
- Sempre que um novo exemplo é encontrado, seus k vizinhos mais próximos dos dados de treinamento são examinados.
- A distância entre dois exemplos pode ser a distância euclidiana entre seus vetores de características.
- A classe majoritária entre os k vizinhos mais próximos é considerada a classe para o exemplo encontrado.