Problemas Clássicos da Computação

Preditores Lineares e Bayesianos

Felipe Augusto Lima Reis
felipe.reis@ifmg.edu.br

INSTITUTO FEDERAL
Minas Gerais

Sumário

Introdução

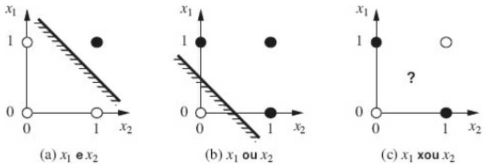
Definição

- Preditores Bayesianos: são aqueles baseados na Regra de Bayes
  - Classificador Ótimo de Bayes;
  - Naïve Bayes;
- Preditores Lineares: são aqueles que realizam separação linear de conjuntos
  - Regressão Linear;
  - Regressão Logística;
  - Perceptron.

Separabilidade Linear

Separabilidade Linear

- Um conjunto é linearmente separável se existir uma hiperplano capaz de separá-lo em duas classes diferentes
  - Em duas dimensões, o hiperplano corresponde a uma reta;
  - Em três dimensões, o hiperplano corresponde a um plano;
- Esse hiperplano é chamado de fronteira de decisão ou função discriminante[1].

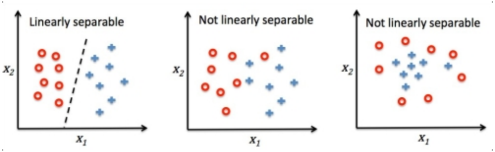


Fonte: [Russel and Norvig, 2013]

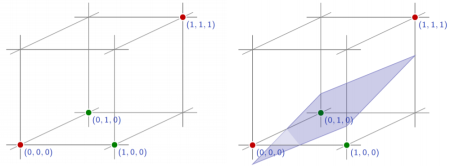
[1] As fronteiras de decisão não são obrigatoriamente lineares.

Separabilidade Linear - Exemplos

- Exemplos de conjuntos, representados em duas e três dimensões, linearmente e não linearmente separáveis.



Fonte: [Raschka, 2015]



Fonte: Francois Fleuret at EPFL apud [Lee, 2020]

Conceitos Básicos de Estatística

Variáveis Independentes e Dependentes

- Variáveis Independentes: são aquelas que são manipuladas ou controladas em um experimento;
- Variáveis Dependentes: são aquelas cujos valores dependem da manipulação da variável independente [Reis, 2006].
  - Possível resultado (ou efeito) do experimento.

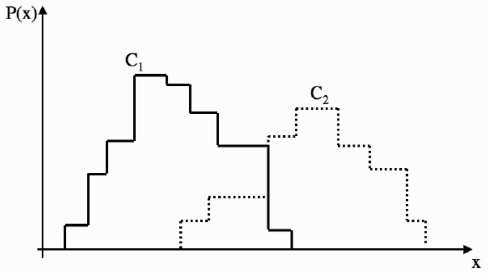
Variáveis Independentes e Dependentes

- Ex. 1: Suponha um experimento no qual deseja-se avaliar a influência da luminosidade sobre o crescimento de planta.
  - Variável independente: luminosidade;
  - Variável dependente: crescimento da planta.
- Ex. 2: Suponha um experimento no qual deseja-se avaliar a influência do *stress* sobre a frequência cardíaca.
  - Variável independente: *stress*;
  - Variável dependente: frequência cardíaca.

Regra de Bayes

Probabilidade Condicional[2].

- Suponha duas classes C_1 e C_2, conforme distribuição abaixo;
  - É possível perceber que as duas classes possuem sobreposição;
  - Os elementos da classe C_1 são mais frequentes que os elementos da classe C_2, ou seja, o conjunto é desbalanceado.



Fonte: [Marsland, 2014]

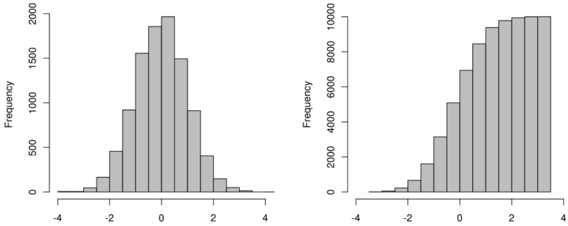
[2] Baseado em [Marsland, 2014]

Probabilidade Condicional

- Suponha uma probabilidade estimada da classe C_1 em 75%;
- Usando esse conhecimento prévio, temos as seguintes probabilidades: $P(C_1) = 0.75$ e $P(C_2) = 0.25$;
- Podemos definir a probabilidade condicional de C_1 para um dado valor x como $X : P(C_1|X)$;

Probabilidade Conjunta

- Para quantizar[3] as medidas de x, podemos ter auxílio de um histograma[4], computando a probabilidade conjunta $P(C_i, X_j)$;
    - Tal probabilidade é equivalente a medida da classe C_i em uma barra X_j do histograma;



Fonte: [Wikipedia contributors, 2020a]

---

[3] Quantificar em valores discretos.

[4] Informações básicas sobre histogramas: `https://pt.wikipedia.org/wiki/Histograma`.

Probabilidade Conjunta

- De forma semelhante, podemos interpretar $P(X_j|C_i)$ como a frequência de uma medida X_j em uma classe C_i;
- Temos, então, duas métricas: a probabilidade conjunta $P(C_i, X_j)$ e a probabilidade condicional $P(X_j|C_i)$
- O link entre as métricas é conhecida como **Regra de Bayes**[5];

[5] Também conhecido como Teorema de Bayes.

Regra de Bayes

- A relação entre probabilidade condicional e conjunta pode ser escrita como:

$$\boxed{P(C_i, X_j) = P(X_j|C_i)P(C_i)} \quad \text{ou} \quad \boxed{P(C_i, X_j) = P(C_i|X_j)P(X_j)}$$

- Reescrevendo as equações anteriores, temos a equação do Teorema de Bayes:

$$\boxed{P(C_i|X_j) = \frac{P(X_j|C_i)P(C_i)}{P(X_j)}}$$

Regra de Bayes

- De forma genérica, podemos definir, para dois eventos distintos A e B:

$$\boxed{P(A|B) = \frac{P(B|A)P(A)}{P(B)}}$$

  - $P(A|B)$: probabilidade condicional de um evento A ocorrer, dado que um evento B é verdadeiro;
  - $P(B|A)$: probabilidade condicional de um evento B ocorrer, dado um evento A é verdadeiro;
  - $P(A)$: probabilidade à priori de um evento A ocorrer;
  - $P(B)$: probabilidade à priori de um evento B ocorrer (usado para normalização das probabilidades).

$P(A)$ e $P(B)$ são conhecidas como probabilidades marginais.

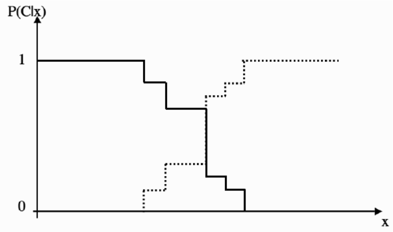
Regra de Bayes

INSTITUTO FEDERAL Minas Gerais

- Segundo [Marsland, 2014], a Regra de Bayes **"é a equação mais importante em Aprendizado de Máquinas"**;
    - Ela relaciona a probabilidade à *posteriori* (ou posterior) $P(C_i|X_j)$ à probabilidade à *priori* $P(C_i)$ e à probabilidade condicional de classe $P(X_j|C_i)$.
- A importância da Regra de Bayes deriva da possibilidade de encontrar uma probabilidade à *posteriori*, a partir de cálculos simples [Marsland, 2014];

Classificador de Bayes

- É possível estimar as probabilidades à *priori*, somente pela análise de frequência no conjunto de treinamento
  - A partir da criação do histograma, é possível gerar as probabilidades condicionais de classe[6].
- A probabilidade à *posteriori* (figura) pode ser usada em novas observações.



Fonte: [Marsland, 2014]

[6] Deve-se ressaltar que, apesar da solução apresentada, a Regra de Bayes não vale apenas para eventos discretos.

Classificador de Bayes

- Um classificador que utilize a abordagem Bayesiana deve escolher a classe de maior valor de probabilidade a *posteriori* [do Patrocínio Jr., 2018].
- Cada nova observação pode ser vinculada a uma classe C_i e um vetor de características x, usando a seguinte fórmula:

$$P(C_i|x) > P(C_j|x) \quad \forall i \neq j$$

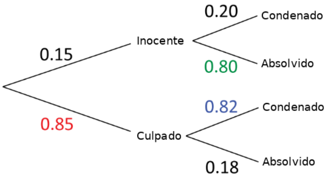
- Essa abordagem é chamada de *Maximum à Posteriori* (MAP) e auxilia na escolha da classe de saída [Marsland, 2014]
  - Um Classificador Ótimo de Bayes é aquele que faz a previsão mais provável para um novo exemplo, desconhecido.

Regra de Bayes - Exemplo [Calcworkshop, 2020]

- Suponha que em um julgamento criminal por júri, tenhamos as seguintes probabilidades:
  - Condenação, caso o réu seja culpado: 82%;
  - Absolvição, caso o réu seja inocente: 80%;
- Suponha que 85% de todos os réus sejam realmente culpados.
- Suponha que um determinado réu seja condenado.
- Encontre a probabilidade dele ser inocente.

Regra de Bayes - Exemplo [Calcworkshop, 2020]

- Primeiramente, podemos criar um diagrama contendo todas as possibilidades:



Fonte: Traduzido de [Calcworkshop, 2020]

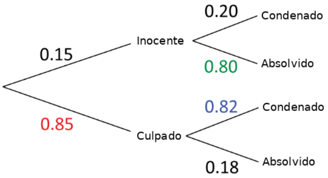
- Definida as condições, podemos calcular a probabilidade:

$$P(inocente|condenado) = \frac{P(condenado|inocente) \times P(inocente)}{P(condenado)}$$

$$\frac{(0.20)(0.15)}{(0.15 \times 0.2) + (0.85 \times 0.82)} = 0.0413$$

Regra de Bayes - Exemplo [Calcworkshop, 2020]

- Primeiramente, podemos criar um diagrama contendo todas as possibilidades:



Fonte: Traduzido de [Calcworkshop, 2020]

- Definida as condições, podemos calcular a probabilidade:

$$\boxed{P(inocente|condenado) = \frac{P(condenado|inocente) \times P(inocente)}{P(condenado)}}$$

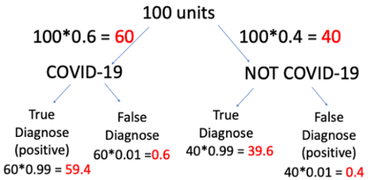
$$\frac{(0.20)(0.15)}{(0.15 \times 0.2) + (0.85 \times 0.82)} = 0.0413$$

Regra de Bayes - Exemplo [Gumusbas, 2020]

- Suponha um teste de COVID-19, para identificação de possíveis doentes, com acurácia de 99%.
- Suponha, ainda, que 60% de todas as pessoas que realizam o testes estão realmente infectadas.
- Se o teste de um paciente for positivo, qual é a probabilidade de ele realmente ter a doença?

Regra de Bayes - Exemplo [Gumusbas, 2020]

- Primeiramente, podemos criar o diagrama correspondente:



Fonte: Traduzido de [Gumusbas, 2020]

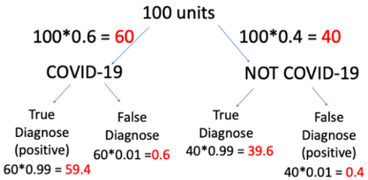
- Definida as condições, podemos calcular a probabilidade:

$$\boxed{P(covid19|positivo) = \frac{P(positivo|covid19) \times P(covid19)}{P(positivo)}}$$

$$\frac{(0.99 \times 0.6)}{(0.6 \times 0.99) + (0.4 \times 0.01)} = 0.9933$$

Regra de Bayes - Exemplo [Gumusbas, 2020]

- Primeiramente, podemos criar o diagrama correspondente:



Fonte: Traduzido de [Gumusbas, 2020]

- Definida as condições, podemos calcular a probabilidade:

$$\boxed{P(covid19|positivo) = \frac{P(positivo|covid19) \times P(covid19)}{P(positivo)}}$$

$$\frac{(0.99 \times 0.6)}{(0.6 \times 0.99) + (0.4 \times 0.01)} = 0.9933$$

Naïve Bayes

Naïve Bayes

- Suponha que tenhamos um vetor de valores de *features*, com muitas características a serem avaliadas;
- O objetivo do classificador é estimar a probabilidade para cada um dos elementos:

$$P(X_j|C_i) = P(X_j^1, X_j^2, ..., X_j^n|C_i)$$

- Suponha que cada uma das *features* sejam independentes uma das outras
  - Devido a essa característica, o classificador é chamado de *naïve* (ingênuo);
  - Na prática, muitas vezes, existe uma correlação implícita entre variáveis - que são desconsideradas pelo algoritmo.

Naïve Bayes

INSTITUTO FEDERAL Minas Gerais

- Suponha a existência n diferentes características, gerando um conjunto $a_1, a_2, ..., a_n$;
- Suponha conjunto de features:

$$h_{bayes} = P(X_j^1 = a_1, X_j^2 = a_2, ..., X_j^n = a_n | C_i)$$

- A probabilidade é dada pela multiplicação de todas as probabilidades individuais:

$$P(X_j^1 = a_1 | C_i) \times P(X_j^2 = a_2 | C_i) \times ... \times P(X_j^n = a_n | C_i) = \prod_k P(X_j^k = a_k | C_i)$$

- O classificador Naïve Bayes selecionará a classe C_i máxima, dada pela seguinte equação: [Marsland, 2014]

$$\boxed{h_{bayes}(x) = (C_i) \prod_k P(X_j^k = a_k | C_i)}$$

Naïve Bayes (Notação Alternativa)

- Notação de [Shalev-Shwartz and Ben-David, 2014]
  - Um Classificador Ótimo de Bayes é definido como:

$$h_{bayes}(x) = \underset{y \in \{0,1\}}{argmax}\ \mathcal{P}[Y = y | X = x]$$

  - O classificador Naïve Bayes selecionará a classe $\mathcal{P}[Y = y|$ de maior distribuição, de acordo com a seguinte equação:

$$\boxed{h_{bayes}(x) = \underset{y \in \{0,1\}}{argmax}\ \mathcal{P}[Y = y] \prod_{i=1}^{d} \mathcal{P}[X_i = x_i | Y = y]}$$

Naïve Bayes

- Apesar do classificador corresponder a uma simplificação, ele possui resultados comparáveis a outros métodos em alguns domínios [Marsland, 2014];
  - Onde as variáveis são realmente independentes umas das outras, o classificador produz exatamente a classificação MAP;
- Para um número de *features* d, define-se a complexidade dos algoritmos: [Shalev-Shwartz and Ben-David, 2014]
  - Ótimo de Bayes: 2^d;
  - Naïve Bayes: $2d + 1$;

Naïve Bayes - Exemplo [Prabhakaran, 2020]

- Suponha um conjunto de 1000 frutas, que devam ser classificadas em "banana", "maracujá" ou "outra fruta";
- A classificação será feita com base em 3 características binárias: comprimento (longa / curta), doçura (doce / não doce) e cor (amarela / outra);
- Os dados foram compilados na seguinte tabela:

Tipo	Longa	Não Longa	Doce	Não Doce	Amarela	Não Amarela	**Total**
Banana	400	100	350	150	450	50	500
Maracujá	0	300	150	150	300	0	300
Outra	100	100	150	50	50	150	200
Total	500	500	650	350	800	200	1000

Adaptado de [Prabhakaran, 2020].

Naïve Bayes - Exemplo [Prabhakaran, 2020]

- Suponha que uma fruta com o seguinte aspecto: longa, doce e amarela
    - A partir das características, a fruta deverá ser classificada.
- Podemos utilizar nessa tarefa um classificador Naïve Bayes
    - Para isso, iremos avaliar, de forma independente, as probabilidades da fruta ser longa, doce e amarela;
    - A classe que tiver maior probabilidade será escolhida.

Naïve Bayes - Exemplo [Prabhakaran, 2020]

- Etapa 1: Cálculo das probabilidades individuais
  1. Calcular probabilidades à priori:
     - P(Y=banana) = 500/1000 = 0.5;
     - P(Y=maracujá) = 300/1000 = 0.3;
     - P(Y=outra) = 200/1000 = 0.2;
  2. Calcular probabilidades de cada uma das características:
     - P(x_1=longa) = 500/1000 = 0.5;
     - P(x_2=doce) = 650/1000 = 0.65;
     - P(x_3=amarela) = 800/1000 = 0.8;

Naïve Bayes - Exemplo [Prabhakaran, 2020]

- Etapa 2: Cálculo das probabilidades condicionais
  1. Banana:
     - P(x_1=longa | Y=banana) = 400/500 = 0.8;
     - P(x_2=doce | Y=banana) = 350/500 = 0.7;
     - P(x_3=amarela | Y=banana) = 450/500 = 0.9;
  2. Maracujá:
     - P(x_1=longa | Y=maracujá) = 0/300 = 0;
     - P(x_2=doce | Y=maracujá) = 150/300 = 0.5;
     - P(x_3=amarela | Y=maracujá) = 300/300 = 1.0;
  3. Outra fruta:
     - P(x_1=longa | Y=outra) = 100/200 = 0.5;
     - P(x_2=doce | Y=outra) = 150/200 = 0.75;
     - P(x_3=amarela | Y=outra) = 50/200 = 0.25;

Naïve Bayes - Exemplo [Prabhakaran, 2020]

- Etapa 3: Cálculo das probabilidades à *posteriori*

$$\boxed{P(Fruta|Caracteristica) = P(Longa|Fruta) \times P(Doce|Fruta) \times P(Amarela|Fruta) \times P(Fruta)}$$

1. Banana - P(Y=banana | x=longa, doce, amarela):
   - $P(A|B) = ((0.8 \times 0.7 \times 0.9) \times 0.5) = 0.252$
2. Maracujá - P(Y=maracujá | x=longa, doce, amarela):
   - $P(A|B) = ((0 \times 0.5 \times 1.0) \times 0.3) = 0$
3. Outra fruta - P(Y=outra | x=longa, doce, amarela):
   - $P(A|B) = ((0.5 \times 0.75 \times 0.25) \times 0.2) = 0.019$

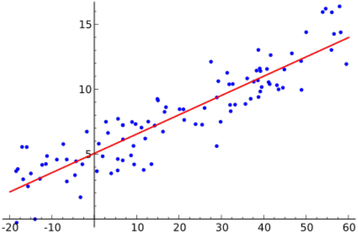
Tipos de Classificadores Naïve Bayes

- Os principais tipos de classificadores Naïve Bayes são: [Metsis et al., 2006] [Shin, 2020]
  - Multinomial: assume que cada valor $P(x_n|y)$ segue uma distribuição multinomial
    - Utilizado em problemas de classificação que contém uma frequência ou repetição de itens;
  - Bernoulli: aqueles em que os preditores são booleanos
    - A distribuição das variáveis utilizada é semelhante ao Naïve Bayes Multinomial;
  - Gaussiano: assumes que os valores contínuos são amostras de uma distribuição gaussiana.

Regressão Linear

Regressão Linear

- A regressão linear é uma ferramenta estatística utilizada para correlacionar uma variável dependente e uma ou mais variáveis explanatórias[7] [Shalev-Shwartz and Ben-David, 2014];
  - Busca aprender uma função linear $h : \mathbb{R}^d \rightarrow \mathbb{R}$ que melhor se aproxima da relação entre as variáveis do problema.



Fonte: [Wikipedia contributors, 2020b]

[7] Também chamadas de variáveis independentes ou preditoras [Ribeiro Jr., 2015].

Regressão Linear

- No contexto de Aprendizado de Máquinas (AM), devem ser separados dos problemas de regressão e classificação
  - Em regressão, ajusta-se um hiperplano aos dados existentes;
  - Em classificação, busca-se um hiperplano que separa as classes [Marsland, 2014].
- Em alguns casos, os problemas de classificação podem ser transformados em regressão
  - Modo 1: adição de uma variável indicadora, que indica a classe ao qual um dado ponto pertence
    - Nesse problema, o objetivo é prever a variável indicadora, o que corresponde a uma regressão;
  - Modo 2: realizar uma regressão repetida para cada uma das classes [Marsland, 2014].

Regressão Linear

- Na regressão, faz-se uma previsão sobre um valor desconhecido $\hat{y}$, por meio de valores conhecidos x_i;
- Como definido previamente, a regressão deseja obter o hiperplano que melhor se ajusta ao conjunto de dados
  - Para isso, busca-se minimizar a distância entre cada ponto e o hiperplano;
  - O objetivo é minimizar a função de erro que mede a soma de todas as distâncias.

Função de Perda

- Para cálculo da perda, é comum o uso da função Squared-loss function[8]: [Marsland, 2014]
  - Corresponde à diferença quadrática entre o valor predito $\hat{y}$ e o valor esperado y;

$$\boxed{\lambda(x) = (\hat{y} - y)^2}$$

- A função de Risco Empírico é chamada de Mean Squared Error (MSE) [Shalev-Shwartz and Ben-David, 2014]
  - Os erro correspondem à soma da perda em todos os pontos;
  - A função MSE é estritamente positiva.

$$\boxed{MSE = \frac{1}{m}\sum_{i=1}^{m}(\hat{y}_i - y_i)^2}$$

[8] Tradução direta: Função de perda quadrática.

Regressão Logística

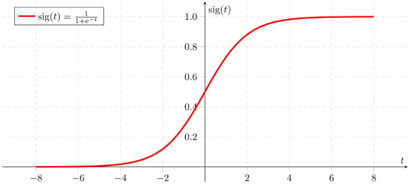
Regressão Logística

- Regressão logística é utilizada para tarefas de classificação, onde é atribuída uma probabilidade entre 0 e 1 a cada classe [Shalev-Shwartz and Ben-David, 2014]
  - Formalmente, o objetivo é aprender uma função $h(x)$ para $\mathbb{R}^d$, no intervalo $[0, 1]$;
  - Em estatística, pode ser utilizada para modelar a probabilidade de um evento ou classe existir;
  - A soma das probabilidades deve ser igual a 1.

Regressão Logística

- A hipótese associada a cada classe da regressão logística é correspondente à composição da função sigmóide sobre classes de funções lineares [Shalev-Shwartz and Ben-David, 2014]
  - Na regressão logística, é utilizada a função logística

$$\boxed{\sigma(t) = \frac{e^t}{e^t + 1} = \frac{1}{1 + e^{-t}}}$$



Fonte: [Wikipedia contributors, 2020c]

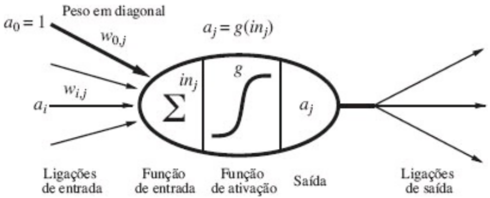
Perceptrons

Neurônios Artificiais

- O modelo de neurônio proposto por McCulloch e Pitts (1943) possuem semelhança com neurônios biológicos [da Silva, 2014]:
  - **Entrada**: sinais $\{x_1, x_2, ..., x_n\}$ oriundos do ambiente externo;
  - **Pesos Sinápticos** (w_i): utilizado para definição de "importância" ou "relevância" das entradas para o neurônio;
  - **Bias** (w_b): entrada extra, utilizada para aumentar o grau de liberdade dos ajustes dos pesos;
  - **Corpo**: soma os produtos das entradas e pesos $(x_i \times w_i)$ com o bias (w_b) e aplica a função de ativação;
  - **Função de Ativação**: controla o comportamento do sinal da saída $f(x)$ de um neurônio limitando o intervalo de valores assumidos.

Funcionamento de um Neurônio Artificial

- Funcionamento de um neurônio:
  1. Cada neurônio possui um sinal de entrada x_i;
  2. Esse neurônio i liga-se a outro neurônio j e é capaz de propagar um valor de ativação x_i;
  3. Cada ligação tem um peso numérico w_i, j associado;



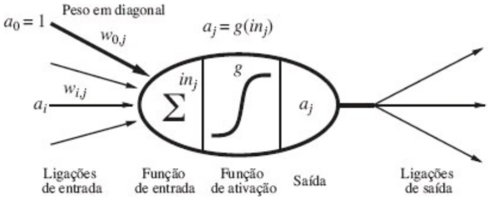
Fonte: [Russel and Norvig, 2013]

Baseado em [da Silva, 2014] e [Russel and Norvig, 2013]

Funcionamento de um Neurônio Artificial

- Funcionamento de um neurônio:
  4. O potencial de ativação de um neurônio é dado pela soma ponderada dos sinais de entrada, somado ao bias (w_b);

$$in_j = \sum_{i=1}^{n} (x_i \cdot w_i) + w_b$$



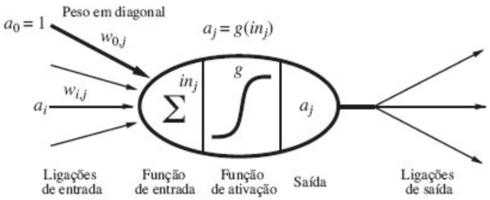
Fonte: [Russel and Norvig, 2013]

Baseado em [da Silva, 2014] e [Russel and Norvig, 2013]

Funcionamento de um Neurônio Artificial

- Funcionamento de um neurônio:
  5. Em seguida, é aplicada uma função de ativação g a essa soma, com objetivo de objetivo de limitar o sinal de saída do neurônio.

$$x_j = g\left(\sum_{i=1}^{n}(x_i \cdot w_i) + w_b\right)$$



Fonte: [Russel and Norvig, 2013]

Baseado em [da Silva, 2014] e [Russel and Norvig, 2013]

Perceptrons

- *Perceptrons* são neurônios simples, propostos por Frank Rosenblatt, em 1958 [Coppin, 2004];
- Esses neurônios são utilizados para classificação de entradas em duas categorias (classificador binário);
  - Indicam se uma entrada pertence a um conjunto ou não;
  - Podem aprender operações booleanas, como AND ou OR;
  - São considerados como um tipo de classificador linear;
  - Introduziram o conceito de treinamento de neurônios e/ou redes neurais.

Perceptrons

- Um Perceptron utiliza uma função degrau (*step*) que retorna $+1$ se a soma de pesos da entrada X for maior que um limiar t e 0 se X é menor ou igual a t [Coppin, 2004]

$$Y = \begin{cases} +1 & for\ X > t \\ 0 & for\ X \leq t \end{cases} \qquad Step(X) = \begin{cases} +1 & for\ X > t \\ 0 & for\ X \leq t \end{cases}$$

Fonte: [Coppin, 2004]

Perceptrons - Processo de Aprendizado

- Para que um Perceptron possa aprender, primeiramente ele deve passar por uma fase de treinamento [Coppin, 2004];
- Nessa fase, os pesos correspondentes às entradas são ajustados
  - Pesos podem ser considerados como indicadores de importância ou relevância das entradas.

Perceptrons - Processo de Aprendizado

- Para início do aprendizado, pesos aleatórios são definidos (em geral, entre -0.5 e +0.5) [Coppin, 2004];
- Uma entrada é dada ao Perceptron, que produz uma saída
  - A saída do Perceptron é comparada com a saída esperada;
  - Caso a saída seja incorreta, os pesos são ajustados e uma nova tentativa é feita;
    - A taxa de ajuste dos pesos é chamada de **taxa de aprendizado**, com valor típico entre 0 e 1;
  - Cada ciclo de ajuste de pesos (tentativa e avaliação de resultados) é chamada de **época**.

Perceptrons - Processo de Aprendizado

- Formalmente, a fórmula de treinamento proposta por Rosenblatt (1960) é dada por [Coppin, 2004];

$$\boxed{w_i \leftarrow w_i + (\eta \times x_i \times e)}$$

  - Onde, η corresponde à taxa de aprendizado (*learning rate*) e e corresponde ao erro/perda (*error*).

- A equação é conhecida como Regra de Treinamento do Perceptron (*perceptron training rule*).

Perceptrons - Processo de Aprendizado

INSTITUTO FEDERAL Minas Gerais

- Considere o treinamento de um perceptron para aprendizado da operação lógica OR.

Epoch	*X1*	*X2*	Expected *Y*	Actual *Y*	Error	*w1*	*w2*
1	0	0	0	0	0	−0.2	0.4
1	0	1	1	1	0	−0.2	0.4
1	**1**	**0**	**1**	**0**	**1**	**0**	**0.4**
1	1	1	1	1	0	0	0.4
2	0	0	0	0	0	0	0.4
2	0	1	1	1	0	0	0.4
2	**1**	**0**	**1**	**0**	**1**	**0.2**	**0.4**
2	1	1	1	1	0	0.2	0.4
3	0	0	0	0	0	0.2	0.4
3	0	1	1	1	0	0.2	0.4
3	1	0	1	1	0	0.2	0.4
3	1	1	1	1	0	0.2	0.4

Fonte: [Coppin, 2004]

Referências I

INSTITUTO FEDERAL Minas Gerais

Calcworkshop (2020).
Bayes theorem - easily explained w/ 7 examples!
Disponível em https://calcworkshop.com/probability/bayes-theorem/.

Coppin, B. (2004).
Artificial Intelligence Illuminated.
Jones and Bartlett illuminated series. Jones and Bartlett Publishers, 1 edition.

da Silva, D. M. (2014).
Inteligência Artificial - Slides de Aula.
IFMG - Instituto Federal de Minas Gerais, Campus Formiga.

do Patrocínio Jr., Z. K. G. (2018).
Aprendizado de Máquina e Reconhecimento de Padrões - Conceitos Básicos / Teoria da Decisão.
Slides de Aula.

Gumusbas, E. (2020).
Bayesâ theorem 101 - example solution.
Disponível em https://towardsdatascience.com/bayes-theorem-101-example-solution-ff54147d6c7f.

Kopec, D. (2019).
Classic Computer Science Problems in Python.
Manning Publications Co, 1 edition.

Lee, S. (2020).
(artificial) neural networks: From perceptron to mlp.
Disponível em: https://github.com/i-Systems/tutorial/blob/gh-pages/KIM/slides/05_ANN.pdf.

Referências II

Marsland, S. (2014).
Machine Learning: An Algorithm Perspective.
CRC Press, 2 edition.
Disponível em: https://homepages.ecs.vuw.ac.nz/ marslast/MLbook.html.

Metsis, V., Androutsopoulos, I., and Paliouras, G. (2006).
Spam filtering with naive bayes - which naive bayes?
In The Third Conference on Email and Anti-Spam.
Disponível em http://www2.aueb.gr/users/ion/docs/ceas2006_paper.pdf.

Prabhakaran, S. (2020).
How naive bayes algorithm works? (with example and full code).
Disponível em https://www.machinelearningplus.com/predictive-modeling/how-naive-bayes-algorithm-works-with-example-and-full-code/.

Raschka, S. (2015).
Python Machine Learning, 1st Edition.
Packt Publishing, 1 edition.

Reis, M. M. (2006).
Conceitos elementares de estatÃstica.
Disponível em: `https://www.inf.ufsc.br/~marcelo.menezes.reis/intro.html`.

Ribeiro Jr., P. J. (2015).
RegressÃ£o.
Disponível em: `http://leg.ufpr.br/~paulojus/CE003/ce003/node9.html`.

Referências III

Richert, W. and Coelho, L. P. (2013).
Building Machine Learning Systems with Python.
Packt Publishing Ltd., 1 edition.

Russel, S. and Norvig, P. (2013).
Inteligência artificial.
Campus - Elsevier, 3 edition.

Shalev-Shwartz, S. and Ben-David, S. (2014).
Understanding Machine Learning: From Theory to Algorithms.
Cambridge University Press, 1 edition.
Disponível em: http://www.cs.huji.ac.il/ shais/UnderstandingMachineLearning.

Shin, T. (2020).
How naive bayes algorithm works? (with example and full code).
Disponível em
https://towardsdatascience.com/a-mathematical-explanation-of-naive-bayes-in-5-minutes-44adebcdb5f8.

Wikipedia contributors (2020a).
Histograma.
Disponível em https://pt.wikipedia.org/wiki/Histograma.

Wikipedia contributors (2020b).
Regression analysis.
Disponível em https://en.wikipedia.org/wiki/Regression_analysis.

Referências IV

Wikipedia contributors (2020c).
Sigmoidfunktion.
Disponível em https://de.wikipedia.org/wiki/Sigmoidfunktion.