

Inteligência Artificial – ACH2016

Aula 17 – Aprendizado Bayesiano

Norton Trevisan Roman
(norton@usp.br)

18 de maio de 2019

Aprendizado

- Até agora, precisávamos de:

Aprendizado

- Até agora, precisávamos de:
 - Modelos bem definidos do problema estudado

Aprendizado

- Até agora, precisávamos de:
 - Modelos bem definidos do problema estudado
 - Conhecimento razoável desse problema

Aprendizado

- Até agora, precisávamos de:
 - Modelos bem definidos do problema estudado
 - Conhecimento razoável desse problema
- Contudo...

Aprendizado

- Até agora, precisávamos de:
 - Modelos bem definidos do problema estudado
 - Conhecimento razoável desse problema
- Contudo...
 - Podemos não ter nem um nem outro

Aprendizado

- Até agora, precisávamos de:
 - Modelos bem definidos do problema estudado
 - Conhecimento razoável desse problema
- Contudo...
 - Podemos não ter nem um nem outro
- Solução

Aprendizado

- Até agora, precisávamos de:
 - Modelos bem definidos do problema estudado
 - Conhecimento razoável desse problema
- Contudo...
 - Podemos não ter nem um nem outro
- Solução
 - Observar vários exemplos

Aprendizado

- Até agora, precisávamos de:
 - Modelos bem definidos do problema estudado
 - Conhecimento razoável desse problema
- Contudo...
 - Podemos não ter nem um nem outro
- Solução
 - Observar vários exemplos
 - Concluir algo sobre eles → aprender

Aprendizado

- Até agora, precisávamos de:
 - Modelos bem definidos do problema estudado
 - Conhecimento razoável desse problema
- Contudo...
 - Podemos não ter nem um nem outro
- Solução
 - Observar vários exemplos
 - Concluir algo sobre eles → aprender
 - Como quando obtínhamos as probabilidades de cada nó na rede bayesiana

Aprendizado

- Quando aprendemos, em geral:

Aprendizado

- Quando aprendemos, em geral:
 - Memorizamos algo

Aprendizado

- Quando aprendemos, em geral:
 - Memorizamos algo
 - Deduzimos fatos a partir de exploração e observação

Aprendizado

- Quando aprendemos, em geral:
 - Memorizamos algo
 - Deduzimos fatos a partir de exploração e observação
 - Melhoramos habilidades pela prática

Aprendizado

- Quando aprendemos, em geral:
 - Memorizamos algo
 - Deduzimos fatos a partir de exploração e observação
 - Melhoramos habilidades pela prática
 - Organizamos novo conhecimento em representações gerais e efetivas

Aprendizado

- Quando aprendemos, em geral:
 - Memorizamos algo
 - Deduzimos fatos a partir de exploração e observação
 - Melhoramos habilidades pela prática
 - Organizamos novo conhecimento em representações gerais e efetivas
- Enfim:
 - Um sistema que aprenda deve obter nova informação de exemplos e melhorar seu desempenho com isso

Aprendizado

- Quando aprendemos, em geral:
 - Memorizamos algo
 - Deduzimos fatos a partir de exploração e observação
 - Melhoramos habilidades pela prática
 - Organizamos novo conhecimento em representações gerais e efetivas
- Enfim:
 - Um sistema que aprenda deve obter nova informação de exemplos e melhorar seu desempenho com isso
 - Corresponde a aprender uma função-alvo que, dada a entrada, indique a saída

Aprendizado

Tipos de Aprendizado

Aprendizado

Tipos de Aprendizado

- Supervisionado

Aprendizado

Tipos de Aprendizado

- Supervisionado
 - A partir de exemplos de pares entrada-saída, aprenda uma função que mapeie da entrada à saída

Aprendizado

Tipos de Aprendizado

- Supervisionado
 - A partir de exemplos de pares entrada-saída, aprenda uma função que mapeie da entrada à saída
 - Necessita de um instrutor

Tipos de Aprendizado

- Supervisionado
 - A partir de exemplos de pares entrada-saída, aprenda uma função que mapeie da entrada à saída
 - Necessita de um instrutor
 - Provém de experimentação

Tipos de Aprendizado

- Supervisionado
 - A partir de exemplos de pares entrada-saída, aprenda uma função que mapeie da entrada à saída
 - Necessita de um instrutor
 - Provém de experimentação
 - Ex: as redes bayesianas vistas

Tipos de Aprendizado

- Supervisionado
 - A partir de exemplos de pares entrada-saída, aprenda uma função que mapeie da entrada à saída
 - Necessita de um instrutor
 - Provém de experimentação
 - Ex: as redes bayesianas vistas
- Não supervisionado

Tipos de Aprendizado

- Supervisionado
 - A partir de exemplos de pares entrada-saída, aprenda uma função que mapeie da entrada à saída
 - Necessita de um instrutor
 - Provém de experimentação
 - Ex: as redes bayesianas vistas
- Não supervisionado
 - Dado apenas o conjunto de entrada, identifique padrões nesse conjunto

Tipos de Aprendizado

- Supervisionado

- A partir de exemplos de pares entrada-saída, aprenda uma função que mapeie da entrada à saída
 - Necessita de um instrutor
 - Provém de experimentação
 - Ex: as redes bayesianas vistas

- Não supervisionado

- Dado apenas o conjunto de entrada, identifique padrões nesse conjunto
- Não sabe exatamente o que aprender → não há um *feedback* explícito

Aprendizado

Tipos de Aprendizado

- Por reforço:

Aprendizado

Tipos de Aprendizado

- Por reforço:
 - O aprendizado se dá por meio de uma recompensa (ou punição)

Aprendizado

Tipos de Aprendizado

- Por reforço:
 - O aprendizado se dá por meio de uma recompensa (ou punição)
 - Só sabe se teve sucesso se, ao final, receber a recompensa

Aprendizado

Tipos de Aprendizado

- Por reforço:
 - O aprendizado se dá por meio de uma recompensa (ou punição)
 - Só sabe se teve sucesso se, ao final, receber a recompensa

Problema maior com aprendizado

- Superajuste (*overfitting*)

Aprendizado

Tipos de Aprendizado

- Por reforço:
 - O aprendizado se dá por meio de uma recompensa (ou punição)
 - Só sabe se teve sucesso se, ao final, receber a recompensa

Problema maior com aprendizado

- Superajuste (*overfitting*)
 - Se o conjunto observado contiver um viés em alguma direção, o sistema pode ficar “viciado”, descartando qualquer hipótese externa a esse conjunto

Aprendizado Supervisionado

Finalidade

Aprendizado Supervisionado

Finalidade

- Classificação

Finalidade

- Classificação
 - Classificar um exemplo novo em um conjunto de classes pré-determinadas

Finalidade

- Classificação
 - Classificar um exemplo novo em um conjunto de classes pré-determinadas
 - Usado para aprender uma função de valores discretos (não necessariamente numéricos)

Finalidade

- Classificação
 - Classificar um exemplo novo em um conjunto de classes pré-determinadas
 - Usado para aprender uma função de valores discretos (não necessariamente numéricos)
 - Ex: Aprovação de crédito – (sim/não)

Finalidade

- Classificação
 - Classificar um exemplo novo em um conjunto de classes pré-determinadas
 - Usado para aprender uma função de valores discretos (não necessariamente numéricos)
 - Ex: Aprovação de crédito – (sim/não)
- Regressão

Finalidade

- Classificação
 - Classificar um exemplo novo em um conjunto de classes pré-determinadas
 - Usado para aprender uma função de valores discretos (não necessariamente numéricos)
 - Ex: Aprovação de crédito – (sim/não)
- Regressão
 - Descobrir o valor associado a um novo exemplo → aprende uma função contínua de valores numéricos

Finalidade

- Classificação
 - Classificar um exemplo novo em um conjunto de classes pré-determinadas
 - Usado para aprender uma função de valores discretos (não necessariamente numéricos)
 - Ex: Aprovação de crédito – (sim/não)
- Regressão
 - Descobrir o valor associado a um novo exemplo → aprende uma função contínua de valores numéricos
 - Ex: Linha de crédito – quantos R\$ podem ser emprestados

Aprendizado Supervisionado

Aprendizado Bayesiano

Aprendizado Supervisionado

Aprendizado Bayesiano

- Puramente probabilístico

Aprendizado Supervisionado

Aprendizado Bayesiano

- Puramente probabilístico
- Fatores:

Aprendizado Bayesiano

- Puramente probabilístico
- Fatores:
 - Cada exemplo aumenta ou reduz a probabilidade de uma hipótese estar correta

Aprendizado Bayesiano

- Puramente probabilístico
- Fatores:
 - Cada exemplo aumenta ou reduz a probabilidade de uma hipótese estar correta
 - Pode usar conhecimento prévio

Aprendizado Bayesiano

- Puramente probabilístico
- Fatores:
 - Cada exemplo aumenta ou reduz a probabilidade de uma hipótese estar correta
 - Pode usar conhecimento prévio
 - Acomoda métodos de predição probabilística

Aprendizado Bayesiano

- Puramente probabilístico
- Fatores:
 - Cada exemplo aumenta ou reduz a probabilidade de uma hipótese estar correta
 - Pode usar conhecimento prévio
 - Acomoda métodos de predição probabilística
 - Novos exemplos são classificados pela combinação da predição de múltiplas hipóteses

Aprendizado Bayesiano

- Puramente probabilístico
- Fatores:
 - Cada exemplo aumenta ou reduz a probabilidade de uma hipótese estar correta
 - Pode usar conhecimento prévio
 - Acomoda métodos de predição probabilística
 - Novos exemplos são classificados pela combinação da predição de múltiplas hipóteses
 - O peso de cada hipótese é sua probabilidade

Aprendizado

Aprendizado Bayesiano

- Baseado em dados e hipóteses

Aprendizado

Aprendizado Bayesiano

- Baseado em dados e hipóteses
 - Dados: evidência

Aprendizado Bayesiano

- Baseado em dados e hipóteses
 - Dados: evidência
 - Hipóteses: possibilidades de classificação

Aprendizado Bayesiano

- Baseado em dados e hipóteses
 - Dados: evidência
 - Hipóteses: possibilidades de classificação
 - Trata-se de um classificador

Aprendizado Bayesiano

- Baseado em dados e hipóteses
 - Dados: evidência
 - Hipóteses: possibilidades de classificação
 - Trata-se de um classificador
- Aprendizado bayesiano calcula a probabilidade de cada hipótese ser a “correta”, dados os dados

Aprendizado Bayesiano

- Baseado em dados e hipóteses
 - Dados: evidência
 - Hipóteses: possibilidades de classificação
 - Trata-se de um classificador
- Aprendizado bayesiano calcula a probabilidade de cada hipótese ser a “correta”, dados os dados
 - $P(H = h_1 | D = d)$

Aprendizado Bayesiano

- Baseado em dados e hipóteses
 - Dados: evidência
 - Hipóteses: possibilidades de classificação
 - Trata-se de um classificador
- Aprendizado bayesiano calcula a probabilidade de cada hipótese ser a “correta”, dados os dados
 - $P(H = h_1 | D = d)$
 - Hipótese “correta” é a que melhor descreve os dados

Aprendizado Bayesiano

- Baseado em dados e hipóteses
 - Dados: evidência
 - Hipóteses: possibilidades de classificação
 - Trata-se de um classificador
- Aprendizado bayesiano calcula a probabilidade de cada hipótese ser a “correta”, dados os dados
 - $P(H = h_1 | D = d)$
 - Hipótese “correta” é a que melhor descreve os dados
 - Usa todas as hipóteses, considerando suas probabilidades, em vez de usar uma única “melhor” hipótese

Aprendizado Bayesiano

- Suponha que queremos saber a probabilidade de fazermos uma determinada observação ($X = x$), dadas as já observadas ($D = d$):

Aprendizado Bayesiano

- Suponha que queremos saber a probabilidade de fazermos uma determinada observação ($X = x$), dadas as já observadas ($D = d$):

$$P(x|d) = \sum_i P(x, h_i|d)$$

Aprendizado Bayesiano

- Suponha que queremos saber a probabilidade de fazermos uma determinada observação ($X = x$), dadas as já observadas ($D = d$):

$$P(x|d) = \sum_i P(x, h_i|d) = \sum_i P(x|d, h_i)P(h_i|d)$$

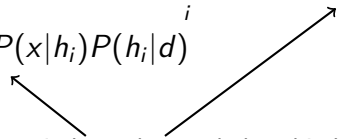
Aprendizado Bayesiano

- Suponha que queremos saber a probabilidade de fazermos uma determinada observação ($X = x$), dadas as já observadas ($D = d$):

$$\begin{aligned} P(x|d) &= \sum_i P(x, h_i|d) = \sum_i P(x|d, h_i)P(h_i|d) \\ &= \sum_i P(x|h_i)P(h_i|d) \end{aligned}$$

Aprendizado Bayesiano

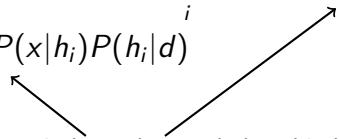
- Suponha que queremos saber a probabilidade de fazermos uma determinada observação ($X = x$), dadas as já observadas ($D = d$):

$$\begin{aligned} P(x|d) &= \sum_i P(x, h_i|d) = \sum_i P(x|d, h_i)P(h_i|d) \\ &= \sum_i P(x|h_i)P(h_i|d) \end{aligned}$$


Assumimos que X e d são independentes dada a hipótese, ou seja, a hipótese é suficiente para determinar a probabilidade de distribuição de X

Aprendizado Bayesiano

- Suponha que queremos saber a probabilidade de fazermos uma determinada observação ($X = x$), dadas as já observadas ($D = d$):

$$\begin{aligned} P(x|d) &= \sum_i P(x, h_i|d) = \sum_i P(x|d, h_i)P(h_i|d) \\ &= \sum_i P(x|h_i)P(h_i|d) \end{aligned}$$


Assumimos que X e d são independentes dada a hipótese, ou seja, a hipótese é suficiente para determinar a probabilidade de distribuição de X

- Temos então que a predição $P(X = x|d)$ é uma média ponderada das predições de cada hipótese individual $P(h_i|d)$

Regra de Bayes

- Lembrando que

$$P(h_i|d) = \frac{P(d|h_i)P(h_i)}{P(d)} \propto P(d|h_i)P(h_i)$$

Regra de Bayes

- Lembrando que

$$P(h_i|d) = \frac{P(d|h_i)P(h_i)}{P(d)} \propto P(d|h_i)P(h_i)$$

- Informalmente, podemos dizer que

$$\textit{posterior} = \frac{\textit{verossimilhança} \times \textit{prévia}}{\textit{evidência}}$$

Regra de Bayes

- Lembrando que

$$P(h_i|d) = \frac{P(d|h_i)P(h_i)}{P(d)} \propto P(d|h_i)P(h_i)$$

Probabilidade posterior

- Informalmente, podemos dizer que

$$\textit{posterior} = \frac{\textit{verossimilhança} \times \textit{prévia}}{\textit{evidência}}$$

Regra de Bayes

- Lembrando que

$$P(h_i|d) = \frac{P(d|h_i)P(h_i)}{P(d)} \propto P(d|h_i)P(h_i)$$

Probabilidade posterior

Verossimilhança (*likelihood*)

- Informalmente, podemos dizer que

$$\textit{posterior} = \frac{\textit{verossimilhança} \times \textit{prévia}}{\textit{evidência}}$$

Aprendizado Bayesiano

Regra de Bayes

- Lembrando que

$$P(h_i|d) = \frac{P(d|h_i)P(h_i)}{P(d)} \propto P(d|h_i)P(h_i)$$

Diagram illustrating the components of Bayes' Rule:

- $P(h_i|d)$ is labeled as **Probabilidade posterior** (Posterior Probability).
- $P(d|h_i)$ is labeled as **Verossimilhança (likelihood)** (Likelihood).
- $P(h_i)$ is labeled as **Probabilidade prévia** (Prior Probability).

- Informalmente, podemos dizer que

$$\textit{posterior} = \frac{\textit{verossimilhança} \times \textit{prévia}}{\textit{evidência}}$$

Aprendizado Bayesiano

Regra de Bayes

$$P(h_i|d) = \frac{P(d|h_i)P(h_i)}{P(d)} \quad \text{posterior} = \frac{\text{verossimilhança} \times \text{prévia}}{\text{evidência}}$$

Aprendizado Bayesiano

Regra de Bayes

$$P(h_i|d) = \frac{P(d|h_i)P(h_i)}{P(d)} \quad \text{posterior} = \frac{\text{verossimilhança} \times \text{prévia}}{\text{evidência}}$$

- $P(d|h_i)$ é a verossimilhança de h_i com respeito a d

Regra de Bayes

$$P(h_i|d) = \frac{P(d|h_i)P(h_i)}{P(d)} \quad \text{posterior} = \frac{\text{verossimilhança} \times \text{prévia}}{\text{evidência}}$$

- $P(d|h_i)$ é a verossimilhança de h_i com respeito a d
- Todo o resto mantido inalterado, indica a categoria h_i para a qual $P(d|h_i)$ é a mais provável de ser verdadeira (ou seja, é máxima)

Regra de Bayes

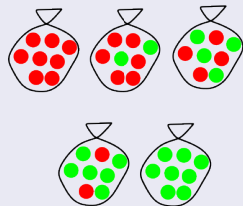
$$P(h_i|d) = \frac{P(d|h_i)P(h_i)}{P(d)} \quad \text{posterior} = \frac{\text{verossimilhança} \times \text{prévia}}{\text{evidência}}$$

- $P(d|h_i)$ é a verossimilhança de h_i com respeito a d
 - Todo o resto mantido inalterado, indica a categoria h_i para a qual $P(d|h_i)$ é a mais provável de ser verdadeira (ou seja, é máxima)
- $P(d)$ pode ser visto meramente como um fator de escala, que garante que as probabilidades posteriores somarão 1

Aprendizado Bayesiano

Exemplo

- Suponha que são fabricados 5 tipos de sacos de balas:

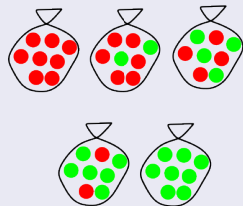


Fonte: Slides de AIMA. Russell & Norvig.

Aprendizado Bayesiano

Exemplo

- Suponha que são fabricados 5 tipos de sacos de balas:
 - h_1 : 100% cereja

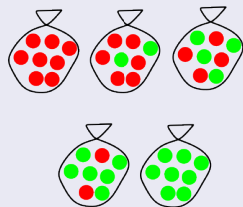


Fonte: Slides de AIMA. Russell & Norvig.

Aprendizado Bayesiano

Exemplo

- Suponha que são fabricados 5 tipos de sacos de balas:
 - h_1 : 100% cereja
 - h_2 : 75% cereja + 25% limão

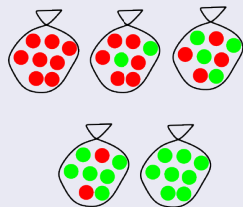


Fonte: Slides de AIMA. Russell & Norvig.

Aprendizado Bayesiano

Exemplo

- Suponha que são fabricados 5 tipos de sacos de balas:
 - h_1 : 100% cereja
 - h_2 : 75% cereja + 25% limão
 - h_3 : 50% cereja + 50% limão

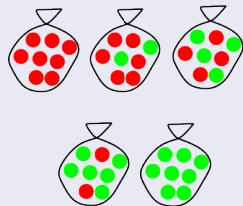


Fonte: Slides de AIMA. Russell & Norvig.

Aprendizado Bayesiano

Exemplo

- Suponha que são fabricados 5 tipos de sacos de balas:
 - h_1 : 100% cereja
 - h_2 : 75% cereja + 25% limão
 - h_3 : 50% cereja + 50% limão
 - h_4 : 25% cereja + 75% limão

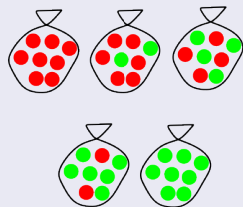


Fonte: Slides de AIMA. Russell & Norvig.

Aprendizado Bayesiano

Exemplo

- Suponha que são fabricados 5 tipos de sacos de balas:
 - h_1 : 100% cereja
 - h_2 : 75% cereja + 25% limão
 - h_3 : 50% cereja + 50% limão
 - h_4 : 25% cereja + 75% limão
 - h_5 : 100% limão

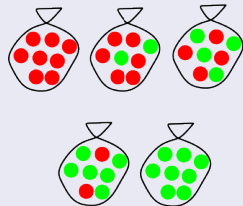


Fonte: Slides de AIMA. Russell & Norvig.

Aprendizado Bayesiano

Exemplo

- Suponha também que esses tipos são fabricados na seguinte proporção:

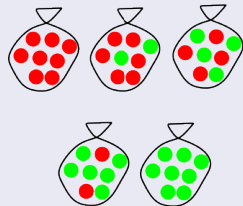


Fonte: Slides de AIMA. Russell & Norvig.

Aprendizado Bayesiano

Exemplo

- Suponha também que esses tipos são fabricados na seguinte proporção:
 - 10% são h_1 (100% cereja)

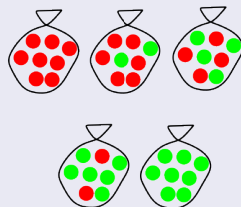


Fonte: Slides de AIMA. Russell & Norvig.

Aprendizado Bayesiano

Exemplo

- Suponha também que esses tipos são fabricados na seguinte proporção:
 - 10% são h_1 (100% cereja)
 - 20% são h_2 (75% cereja + 25% limão)

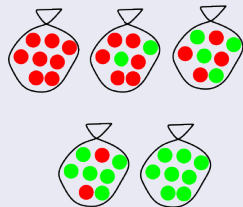


Fonte: Slides de AIMA. Russell & Norvig.

Aprendizado Bayesiano

Exemplo

- Suponha também que esses tipos são fabricados na seguinte proporção:
 - 10% são h_1 (100% cereja)
 - 20% são h_2 (75% cereja + 25% limão)
 - 40% são h_3 (50% cereja + 50% limão)

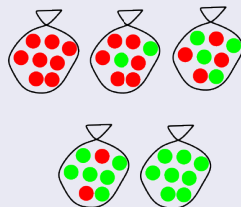


Fonte: Slides de AIMA. Russell & Norvig.

Aprendizado Bayesiano

Exemplo

- Suponha também que esses tipos são fabricados na seguinte proporção:
 - 10% são h_1 (100% cereja)
 - 20% são h_2 (75% cereja + 25% limão)
 - 40% são h_3 (50% cereja + 50% limão)
 - 20% são h_4 (25% cereja + 75% limão)

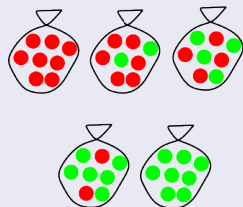


Fonte: Slides de AIMA. Russell & Norvig.

Aprendizado Bayesiano

Exemplo

- Suponha também que esses tipos são fabricados na seguinte proporção:
 - 10% são h_1 (100% cereja)
 - 20% são h_2 (75% cereja + 25% limão)
 - 40% são h_3 (50% cereja + 50% limão)
 - 20% são h_4 (25% cereja + 75% limão)
 - 10% são h_5 (100% limão)



Fonte: Slides de AIMA. Russell & Norvig.

Exemplo

- Suponha agora que você comprou um saco de balas sem saber de que tipo

Exemplo

- Suponha agora que você comprou um saco de balas sem saber de que tipo
- E aleatoriamente retira n balas desse saco, obtendo :



Aprendizado Bayesiano

Exemplo

- Suponha agora que você comprou um saco de balas sem saber de que tipo
 - E aleatoriamente retira n balas desse saco, obtendo :



- Que tipo de saco é esse?

Exemplo

- Suponha agora que você comprou um saco de balas sem saber de que tipo
- E aleatoriamente retira n balas desse saco, obtendo :



- Que tipo de saco é esse?
- Supondo que as observações (os dados) são independentes:

$$P(d|h_i) = \prod_j P(d_j|h_i)$$

onde $d = \{d_1, \dots, d_n\}$ são as n balas já retiradas

Exemplo

- Qual então a probabilidade de observarmos a primeira bala de limão?

Exemplo

- Qual então a probabilidade de observarmos a primeira bala de limão?
- Não há dados anteriores \Rightarrow probabilidade prévia:

Exemplo

- Qual então a probabilidade de observarmos a primeira bala de limão?
- Não há dados anteriores \Rightarrow probabilidade prévia:

$$P(I) = \sum_{i=1}^5 P(I, h_i) = \sum_{i=1}^5 P(I|h_i)P(h_i)$$

Exemplo

- Qual então a probabilidade de observarmos a primeira bala de limão?
- Não há dados anteriores \Rightarrow probabilidade prévia:

$$P(I) = \sum_{i=1}^5 P(I, h_i) = \sum_{i=1}^5 P(I|h_i)P(h_i)$$

$$P(I) = 0 \times 0,1 + 0,25 \times 0,2 + 0,5 \times 0,4 + 0,75 \times 0,2 + 1 \times 0,1 = 0,5$$

Exemplo

- Qual então a probabilidade de observarmos a primeira bala de limão?

- Não há dados anteriores \Rightarrow probabilidade prévia:

$$P(I) = \sum_{i=1}^5 P(I, h_i) = \sum_{i=1}^5 P(I|h_i)P(h_i)$$

$$P(I) = 0 \times 0,1 + 0,25 \times 0,2 + 0,5 \times 0,4 + 0,75 \times 0,2 + 1 \times 0,1 = 0,5$$

- E de observarmos a segunda bala de limão, dada a primeira?

Exemplo

- Qual então a probabilidade de observarmos a primeira bala de limão?
- Não há dados anteriores \Rightarrow probabilidade prévia:

$$P(I) = \sum_{i=1}^5 P(I, h_i) = \sum_{i=1}^5 P(I|h_i)P(h_i)$$

$$P(I) = 0 \times 0,1 + 0,25 \times 0,2 + 0,5 \times 0,4 + 0,75 \times 0,2 + 1 \times 0,1 = 0,5$$

- E de observarmos a segunda bala de limão, dada a primeira?

$$P(I|d) = P(I|\{I\}) = \sum_{i=1}^5 P(I, h_i|\{I\})$$

Exemplo

$$\Rightarrow P(I|d) = \sum_{i=1}^5 P(I|\{I\}, h_i)P(h_i|\{I\})$$

Exemplo

$$\begin{aligned}\Rightarrow P(I|d) &= \sum_{i=1}^5 P(I|\{I\}, h_i)P(h_i|\{I\}) \\ &= \sum_{i=1}^5 P(I|h_i)P(h_i|\{I\})\end{aligned}$$

Aprendizado Bayesiano

Exemplo

$$\begin{aligned}\Rightarrow P(I|d) &= \sum_{i=1}^5 P(I|\{I\}, h_i)P(h_i|\{I\}) \\ &= \sum_{i=1}^5 P(I|h_i)P(h_i|\{I\})\end{aligned}$$

(supondo independência condicional entre I e $d = \{I\}$, dada a hipótese h_i)

Aprendizado Bayesiano

Exemplo

$$\begin{aligned}\Rightarrow P(I|d) &= \sum_{i=1}^5 P(I|\{I\}, h_i)P(h_i|\{I\}) \\ &= \sum_{i=1}^5 P(I|h_i)P(h_i|\{I\})\end{aligned}$$

(supondo independência condicional entre I e $d = \{I\}$, dada a hipótese h_i)

Como $P(h_i|\{I\}) = \frac{P(\{I\}|h_i)P(h_i)}{P(\{I\})}$, então

$$P(I|d) = \frac{1}{P(\{I\})} \sum_{i=1}^5 P(I|h_i)P(\{I\}|h_i)P(h_i)$$

Exemplo

E como $P(d|h_i) = \prod_j P(d_j|h_i) \Rightarrow P(\{I\}|h_i) = P(I|h_i)$, então

$$P(I|d) = \frac{1}{P(\{I\})} \sum_{i=1}^5 P(I|h_i)P(I|h_i)P(h_i)$$

Aprendizado Bayesiano

Exemplo

E como $P(d|h_i) = \prod_j P(d_j|h_i) \Rightarrow P(\{I\}|h_i) = P(I|h_i)$, então

$$\begin{aligned} P(I|d) &= \frac{1}{P(\{I\})} \sum_{i=1}^5 P(I|h_i)P(I|h_i)P(h_i) \\ &= \frac{1}{P(\{I\})} \sum_{i=1}^5 P(I|h_i)^2 P(h_i) \end{aligned}$$

Aprendizado Bayesiano

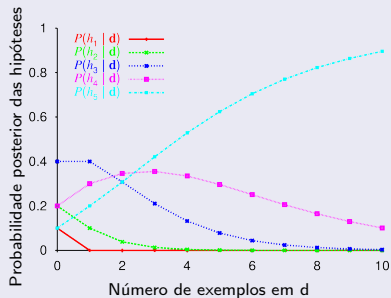
Exemplo

E como $P(d|h_i) = \prod_j P(d_j|h_i) \Rightarrow P(\{I\}|h_i) = P(I|h_i)$, então

$$\begin{aligned} P(I|d) &= \frac{1}{P(\{I\})} \sum_{i=1}^5 P(I|h_i)P(I|h_i)P(h_i) \\ &= \frac{1}{P(\{I\})} \sum_{i=1}^5 P(I|h_i)^2 P(h_i) \\ &= \frac{0,325}{P(\{I\})} = \frac{0,325}{P(I)} = \frac{0.325}{0,5} = 0,65 \end{aligned}$$

Aprendizado Bayesiano

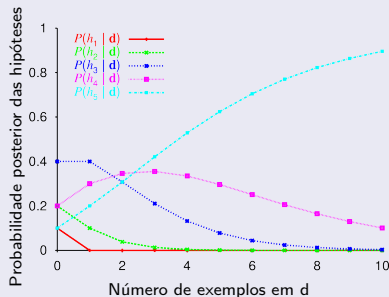
Exemplo



Fonte: Adaptado de slides de AIMA. Russell & Norvig.

Aprendizado Bayesiano

Exemplo

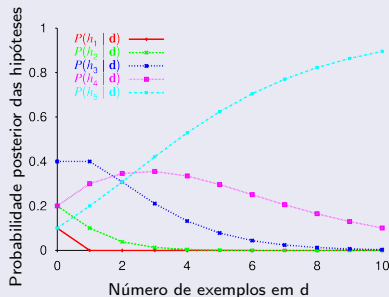


À medida em que vemos mais exemplos, a hipótese real acaba dominando

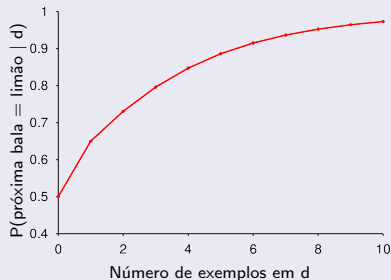
Fonte: Adaptado de slides de AIMA. Russell & Norvig.

Aprendizado Bayesiano

Exemplo



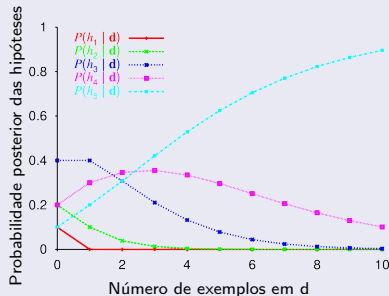
À medida em que vemos mais exemplos, a hipótese real acaba dominando



Fonte: Adaptado de slides de AIMA. Russell & Norvig.

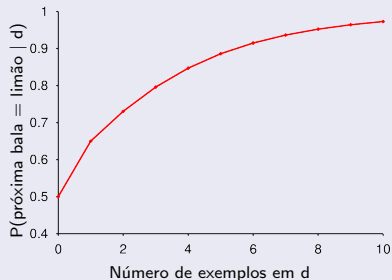
Aprendizado Bayesiano

Exemplo



Naturalmente tende a 1, refletindo o aumento de nossa confiança de que a próxima bola continuará a ser de limão

À medida em que vemos mais exemplos, a hipótese real acaba dominando



Fonte: Adaptado de slides de AIMA. Russell & Norvig.

Aprendizado Bayesiano

Problema

- Calcular $P(x|d)$ pode tornar-se intratável

Aprendizado Bayesiano

Problema

- Calcular $P(x|d)$ pode tornar-se intratável

Solução

- Uma alternativa comum é lidar apenas com a hipótese mais provável, tendo observado o conjunto de dados d

Aprendizado Bayesiano

Problema

- Calcular $P(x|d)$ pode tornar-se intratável

Solução

- Uma alternativa comum é lidar apenas com a hipótese mais provável, tendo observado o conjunto de dados d
 - h_i que maximiza $P(h_i|d)$

Aprendizado Bayesiano

Problema

- Calcular $P(x|d)$ pode tornar-se intratável

Solução

- Uma alternativa comum é lidar apenas com a hipótese mais provável, tendo observado o conjunto de dados d
 - h_i que maximiza $P(h_i|d)$
 - Maximum a Posteriori: $h_{MAP} = \underset{i}{\operatorname{argmax}} P(h_i|d)$

Aprendizado Bayesiano

Problema

- Calcular $P(x|d)$ pode tornar-se intratável

Solução

- Uma alternativa comum é lidar apenas com a hipótese mais provável, tendo observado o conjunto de dados d
 - h_i que maximiza $P(h_i|d)$
 - Maximum a Posteriori: $h_{MAP} = \underset{i}{\operatorname{argmax}} P(h_i|d)$
 - Base para o Naïve Bayes Classifier – **Classificador Bayesiano Ingênuo**

Classificador Bayesiano Ingênuo

Classificador Bayesiano Ingênuo

Aplicação

Classificador Bayesiano Ingênuo

Aplicação

- Tarefas de aprendizado onde:

Classificador Bayesiano Ingênuo

Aplicação

- Tarefas de aprendizado onde:
 - Cada instância do problema pode ser descrita por uma conjunção de atributos

Classificador Bayesiano Ingênuo

Aplicação

- Tarefas de aprendizado onde:
 - Cada instância do problema pode ser descrita por uma conjunção de atributos
 - A função-alvo pode retornar qualquer valor de algum conjunto finito

Classificador Bayesiano Ingênuo

Aplicação

- Tarefas de aprendizado onde:
 - Cada instância do problema pode ser descrita por uma conjunção de atributos
 - A função-alvo pode retornar qualquer valor de algum conjunto finito

Função-alvo:

- Uma função matemática que, para cada instância do problema, indique a categoria à qual ela pertence

Classificador Bayesiano Ingênuo

Aplicação

- Tarefas de aprendizado onde:
 - Cada instância do problema pode ser descrita por uma conjunção de atributos
 - A função-alvo pode retornar qualquer valor de algum conjunto finito

Função-alvo:

- Uma função matemática que, para cada instância do problema, indique a categoria à qual ela pertence
 - É a função efetivamente aprendida

Classificador Bayesiano Ingênuo

- Funcionamento básico:

Classificador Bayesiano Ingênuo

- Funcionamento básico:
 - Com base na evidência, e usando a regra de Bayes, atualiza a probabilidade de cada hipótese

Classificador Bayesiano Ingênuo

- Funcionamento básico:
 - Com base na evidência, e usando a regra de Bayes, atualiza a probabilidade de cada hipótese
 - Após incorporar a evidência, escolhe a hipótese mais provável, para cada novo dado que precise classificar

Classificador Bayesiano Ingênuo

- Funcionamento básico:
 - Com base na evidência, e usando a regra de Bayes, atualiza a probabilidade de cada hipótese
 - Após incorporar a evidência, escolhe a hipótese mais provável, para cada novo dado que precise classificar
- Alternativamente:

Classificador Bayesiano Ingênuo

- Funcionamento básico:
 - Com base na evidência, e usando a regra de Bayes, atualiza a probabilidade de cada hipótese
 - Após incorporar a evidência, escolhe a hipótese mais provável, para cada novo dado que precise classificar
- Alternativamente:
 - São dados exemplos de treino da função-alvo

Classificador Bayesiano Ingênuo

- Funcionamento básico:
 - Com base na evidência, e usando a regra de Bayes, atualiza a probabilidade de cada hipótese
 - Após incorporar a evidência, escolhe a hipótese mais provável, para cada novo dado que precise classificar
- Alternativamente:
 - São dados exemplos de treino da função-alvo
 - Uma nova instância é apresentada $X = \langle a_1, a_2, \dots, a_n \rangle$

Classificador Bayesiano Ingênuo

- Funcionamento básico:
 - Com base na evidência, e usando a regra de Bayes, atualiza a probabilidade de cada hipótese
 - Após incorporar a evidência, escolhe a hipótese mais provável, para cada novo dado que precise classificar
- Alternativamente:
 - São dados exemplos de treino da função-alvo
 - Uma nova instância é apresentada $X = \langle a_1, a_2, \dots, a_n \rangle$
 - O classificador calcula o valor da função-alvo para a nova instância

Classificador Bayesiano Ingênuo

Para classificar uma nova instância:

Classificador Bayesiano Ingênuo

Para classificar uma nova instância:

- A associa à hipótese mais provável, dados os atributos que descrevem a instância

Classificador Bayesiano Ingênuo

Para classificar uma nova instância:

- A associa à hipótese mais provável, dados os atributos que descrevem a instância

$$h_{MAP} = \underset{i}{\operatorname{argmax}} P(h_i | a_1, a_2, \dots, a_n)$$

Classificador Bayesiano Ingênuo

Para classificar uma nova instância:

- A associa à hipótese mais provável, dados os atributos que descrevem a instância

$$h_{MAP} = \operatorname{argmax}_i P(h_i | a_1, a_2, \dots, a_n)$$

- Então:

$$h_{MAP} = \operatorname{argmax}_i \frac{P(a_1, a_2, \dots, a_n | h_i) P(h_i)}{P(a_1, a_2, \dots, a_n)}$$

Classificador Bayesiano Ingênuo

Para classificar uma nova instância:

- A associa à hipótese mais provável, dados os atributos que descrevem a instância

$$h_{MAP} = \underset{i}{\operatorname{argmax}} P(h_i | a_1, a_2, \dots, a_n)$$

- Então:

$$h_{MAP} = \underset{i}{\operatorname{argmax}} \frac{P(a_1, a_2, \dots, a_n | h_i) P(h_i)}{P(a_1, a_2, \dots, a_n)}$$



$$h_{MAP} = \underset{i}{\operatorname{argmax}} P(a_1, a_2, \dots, a_n | h_i) P(h_i)$$

Classificador Bayesiano Ingênuo

Para classificar uma nova instância:

- A associa à hipótese mais provável, dados os atributos que descrevem a instância

$$h_{MAP} = \operatorname{argmax}_i P(h_i | a_1, a_2, \dots, a_n)$$

- Então:

$$h_{MAP} = \operatorname{argmax}_i \frac{P(a_1, a_2, \dots, a_n | h_i) P(h_i)}{P(a_1, a_2, \dots, a_n)}$$



$$h_{MAP} = \operatorname{argmax}_i P(a_1, a_2, \dots, a_n | h_i) P(h_i)$$

Por ser constante,
 $P(a_1, a_2, \dots, a_n)$ não
influencia na deter-
minação do máximo

Classificador Bayesiano Ingênuo

- $P(h_i) \rightarrow$ contamos a frequência com que h_i ocorre nos dados de treino

Classificador Bayesiano Ingênuo

- $P(h_i) \rightarrow$ contamos a frequência com que h_i ocorre nos dados de treino
- E $P(a_1, a_2, \dots, a_n | h_i)$?

Classificador Bayesiano Ingênuo

- $P(h_i) \rightarrow$ contamos a frequência com que h_i ocorre nos dados de treino
- E $P(a_1, a_2, \dots, a_n | h_i)$?
 - Difícil, a menos que tenhamos um conjunto de treino realmente grande

Classificador Bayesiano Ingênuo

- $P(h_i) \rightarrow$ contamos a frequência com que h_i ocorre nos dados de treino
- E $P(a_1, a_2, \dots, a_n | h_i)$?
 - Difícil, a menos que tenhamos um conjunto de treino realmente grande
 - Teríamos que ter contagens suficientes para cada valor possível dos atributos, em cada instância

Classificador Bayesiano Ingênuo

- $P(h_i) \rightarrow$ contamos a frequência com que h_i ocorre nos dados de treino
- E $P(a_1, a_2, \dots, a_n | h_i)$?
 - Difícil, a menos que tenhamos um conjunto de treino realmente grande
 - Teríamos que ter contagens suficientes para cada valor possível dos atributos, em cada instância
 - Note que o mesmo par $\langle a_j, h_i \rangle$ aparece várias vezes (conforme a combinação dos demais atributos)

Classificador Bayesiano Ingênuo

- Solução: a parte “ingênua” do classificador

Classificador Bayesiano Ingênuo

- Solução: a parte “ingênua” do classificador
 - Supõe que os atributos são condicionalmente independentes, dado h_i

Classificador Bayesiano Ingênuo

- Solução: a parte “ingênua” do classificador
 - Supõe que os atributos são condicionalmente independentes, dado h_i

$$P(a_1, a_2, \dots, a_n | h_i) = \prod_{j=1}^n P(a_j | h_i)$$

Classificador Bayesiano Ingênuo

- Solução: a parte “ingênua” do classificador
 - Supõe que os atributos são condicionalmente independentes, dado h_i

$$P(a_1, a_2, \dots, a_n | h_i) = \prod_{j=1}^n P(a_j | h_i)$$



$$h_{NB} = \operatorname{argmax}_i P(h_i) \prod_{j=1}^n P(a_j | h_i)$$

Classificador Bayesiano Ingênuo

- Solução: a parte “ingênua” do classificador
 - Supõe que os atributos são condicionalmente independentes, dado h_i

$$P(a_1, a_2, \dots, a_n | h_i) = \prod_{j=1}^n P(a_j | h_i)$$

Hipótese “escolhida” pelo classificador para a nova observação



$$h_{NB} = \underset{i}{\operatorname{argmax}} P(h_i) \prod_{j=1}^n P(a_j | h_i)$$

Classificador Bayesiano Ingênuo

- Solução: a parte “ingênua” do classificador
 - Supõe que os atributos são condicionalmente independentes, dado h_i

$$P(a_1, a_2, \dots, a_n | h_i) = \prod_{j=1}^n P(a_j | h_i)$$

Hipótese “escolhida” pelo classificador para a nova observação



Embora possa vir a ser uma boa aproximação, nem sempre é o caso dos atributos serem condicionalmente independentes

$$h_{NB} = \underset{i}{\operatorname{argmax}} P(h_i) \prod_{j=1}^n P(a_j | h_i)$$

Classificador Bayesiano Ingênuo

- Solução: a parte “ingênua” do classificador
 - Supõe que os atributos são condicionalmente independentes, dado h_i

$$P(a_1, a_2, \dots, a_n | h_i) = \prod_{j=1}^n P(a_j | h_i)$$

Hipótese “escolhida” pelo classificador para a nova observação



Embora possa vir a ser uma boa aproximação, nem sempre é o caso dos atributos serem condicionalmente independentes

$$h_{NB} = \operatorname{argmax}_i P(h_i) \prod_{j=1}^n P(a_j | h_i)$$

- Note que o mesmo par $\langle a_j, h_i \rangle$ aparece uma única vez

Classificador Bayesiano Ingênuo

Exemplo

A_1	A_2	A_3	A_4	H
0	1	1	0	1
0	0	1	1	1
1	0	1	0	1
0	0	1	1	1
0	0	0	0	1
1	0	0	1	0
1	1	0	1	0
1	0	0	0	0
1	1	0	1	0
1	0	1	1	0

Classificador Bayesiano Ingênuo

Exemplo

A_1	A_2	A_3	A_4	H
0	1	1	0	1
0	0	1	1	1
1	0	1	0	1
0	0	1	1	1
0	0	0	0	1
1	0	0	1	0
1	1	0	1	0
1	0	0	0	0
1	1	0	1	0
1	0	1	1	0

- $H = \{1, 0\} \rightarrow h_1 = 1, h_2 = 0$

Classificador Bayesiano Ingênuo

Exemplo

A_1	A_2	A_3	A_4	H
0	1	1	0	1
0	0	1	1	1
1	0	1	0	1
0	0	1	1	1
0	0	0	0	1
1	0	0	1	0
1	1	0	1	0
1	0	0	0	0
1	1	0	1	0
1	0	1	1	0

- $H = \{1, 0\} \rightarrow h_1 = 1, h_2 = 0$
- $P(h_1) = 5/10 = 0,5; P(h_2) = 5/10 = 0,5$

Classificador Bayesiano Ingênuo

Exemplo

A_1	A_2	A_3	A_4	H
0	1	1	0	1
0	0	1	1	1
1	0	1	0	1
0	0	1	1	1
0	0	0	0	1
1	0	0	1	0
1	1	0	1	0
1	0	0	0	0
1	1	0	1	0
1	0	1	1	0

- $H = \{1, 0\} \rightarrow h_1 = 1, h_2 = 0$
- $P(h_1) = 5/10 = 0,5; P(h_2) = 5/10 = 0,5$
- $P(A_1 = 0|h_1) = 4/5 = 0,8; P(A_1 = 1|h_1) = 1/5 = 0,2$

Classificador Bayesiano Ingênuo

Exemplo

A_1	A_2	A_3	A_4	H
0	1	1	0	1
0	0	1	1	1
1	0	1	0	1
0	0	1	1	1
0	0	0	0	1
1	0	0	1	0
1	1	0	1	0
1	0	0	0	0
1	1	0	1	0
1	0	1	1	0

- $H = \{1, 0\} \rightarrow h_1 = 1, h_2 = 0$
- $P(h_1) = 5/10 = 0,5; P(h_2) = 5/10 = 0,5$
- $P(A_1 = 0|h_1) = 4/5 = 0,8; P(A_1 = 1|h_1) = 1/5 = 0,2$
- $P(A_2 = 0|h_1) = 4/5 = 0,8; P(A_2 = 1|h_1) = 1/5 = 0,2$

Classificador Bayesiano Ingênuo

Exemplo

A_1	A_2	A_3	A_4	H
0	1	1	0	1
0	0	1	1	1
1	0	1	0	1
0	0	1	1	1
0	0	0	0	1
1	0	0	1	0
1	1	0	1	0
1	0	0	0	0
1	1	0	1	0
1	0	1	1	0

- $H = \{1, 0\} \rightarrow h_1 = 1, h_2 = 0$
- $P(h_1) = 5/10 = 0,5; P(h_2) = 5/10 = 0,5$
- $P(A_1 = 0|h_1) = 4/5 = 0,8; P(A_1 = 1|h_1) = 1/5 = 0,2$
- $P(A_2 = 0|h_1) = 4/5 = 0,8; P(A_2 = 1|h_1) = 1/5 = 0,2$
- $P(A_3 = 0|h_1) = 1/5 = 0,2; P(A_3 = 1|h_1) = 4/5 = 0,8$

Classificador Bayesiano Ingênuo

Exemplo

A_1	A_2	A_3	A_4	H
0	1	1	0	1
0	0	1	1	1
1	0	1	0	1
0	0	1	1	1
0	0	0	0	1
1	0	0	1	0
1	1	0	1	0
1	0	0	0	0
1	1	0	1	0
1	0	1	1	0

- $H = \{1, 0\} \rightarrow h_1 = 1, h_2 = 0$
- $P(h_1) = 5/10 = 0,5; P(h_2) = 5/10 = 0,5$
- $P(A_1 = 0|h_1) = 4/5 = 0,8; P(A_1 = 1|h_1) = 1/5 = 0,2$
- $P(A_2 = 0|h_1) = 4/5 = 0,8; P(A_2 = 1|h_1) = 1/5 = 0,2$
- $P(A_3 = 0|h_1) = 1/5 = 0,2; P(A_3 = 1|h_1) = 4/5 = 0,8$
- $P(A_4 = 0|h_1) = 3/5 = 0,6; P(A_4 = 1|h_1) = 2/5 = 0,4$

Classificador Bayesiano Ingênuo

Exemplo

A_1	A_2	A_3	A_4	H
0	1	1	0	1
0	0	1	1	1
1	0	1	0	1
0	0	1	1	1
0	0	0	0	1
1	0	0	1	0
1	1	0	1	0
1	0	0	0	0
1	1	0	1	0
1	0	1	1	0

- $H = \{1, 0\} \rightarrow h_1 = 1, h_2 = 0$
- $P(h_1) = 5/10 = 0,5; P(h_2) = 5/10 = 0,5$
- $P(A_1 = 0|h_1) = 4/5 = 0,8; P(A_1 = 1|h_1) = 1/5 = 0,2$
- $P(A_2 = 0|h_1) = 4/5 = 0,8; P(A_2 = 1|h_1) = 1/5 = 0,2$
- $P(A_3 = 0|h_1) = 1/5 = 0,2; P(A_3 = 1|h_1) = 4/5 = 0,8$
- $P(A_4 = 0|h_1) = 3/5 = 0,6; P(A_4 = 1|h_1) = 2/5 = 0,4$
- $P(A_1 = 0|h_2) = 0/5 = 0,0; P(A_1 = 1|h_2) = 5/5 = 1,0$

Classificador Bayesiano Ingênuo

Exemplo

A_1	A_2	A_3	A_4	H
0	1	1	0	1
0	0	1	1	1
1	0	1	0	1
0	0	1	1	1
0	0	0	0	1
1	0	0	1	0
1	1	0	1	0
1	0	0	0	0
1	1	0	1	0
1	0	1	1	0

- $H = \{1, 0\} \rightarrow h_1 = 1, h_2 = 0$
- $P(h_1) = 5/10 = 0,5; P(h_2) = 5/10 = 0,5$
- $P(A_1 = 0|h_1) = 4/5 = 0,8; P(A_1 = 1|h_1) = 1/5 = 0,2$
- $P(A_2 = 0|h_1) = 4/5 = 0,8; P(A_2 = 1|h_1) = 1/5 = 0,2$
- $P(A_3 = 0|h_1) = 1/5 = 0,2; P(A_3 = 1|h_1) = 4/5 = 0,8$
- $P(A_4 = 0|h_1) = 3/5 = 0,6; P(A_4 = 1|h_1) = 2/5 = 0,4$
- $P(A_1 = 0|h_2) = 0/5 = 0,0; P(A_1 = 1|h_2) = 5/5 = 1,0$
- $P(A_2 = 0|h_2) = 3/5 = 0,6; P(A_2 = 1|h_2) = 2/5 = 0,4$

Classificador Bayesiano Ingênuo

Exemplo

A_1	A_2	A_3	A_4	H
0	1	1	0	1
0	0	1	1	1
1	0	1	0	1
0	0	1	1	1
0	0	0	0	1
1	0	0	1	0
1	1	0	1	0
1	0	0	0	0
1	1	0	1	0
1	0	1	1	0

- $H = \{1, 0\} \rightarrow h_1 = 1, h_2 = 0$
- $P(h_1) = 5/10 = 0,5; P(h_2) = 5/10 = 0,5$
- $P(A_1 = 0|h_1) = 4/5 = 0,8; P(A_1 = 1|h_1) = 1/5 = 0,2$
- $P(A_2 = 0|h_1) = 4/5 = 0,8; P(A_2 = 1|h_1) = 1/5 = 0,2$
- $P(A_3 = 0|h_1) = 1/5 = 0,2; P(A_3 = 1|h_1) = 4/5 = 0,8$
- $P(A_4 = 0|h_1) = 3/5 = 0,6; P(A_4 = 1|h_1) = 2/5 = 0,4$
- $P(A_1 = 0|h_2) = 0/5 = 0,0; P(A_1 = 1|h_2) = 5/5 = 1,0$
- $P(A_2 = 0|h_2) = 3/5 = 0,6; P(A_2 = 1|h_2) = 2/5 = 0,4$
- $P(A_3 = 0|h_2) = 4/5 = 0,8; P(A_3 = 1|h_2) = 1/5 = 0,2$

Classificador Bayesiano Ingênuo

Exemplo

A_1	A_2	A_3	A_4	H
0	1	1	0	1
0	0	1	1	1
1	0	1	0	1
0	0	1	1	1
0	0	0	0	1
1	0	0	1	0
1	1	0	1	0
1	0	0	0	0
1	1	0	1	0
1	0	1	1	0

- $H = \{1, 0\} \rightarrow h_1 = 1, h_2 = 0$
- $P(h_1) = 5/10 = 0,5; P(h_2) = 5/10 = 0,5$
- $P(A_1 = 0|h_1) = 4/5 = 0,8; P(A_1 = 1|h_1) = 1/5 = 0,2$
- $P(A_2 = 0|h_1) = 4/5 = 0,8; P(A_2 = 1|h_1) = 1/5 = 0,2$
- $P(A_3 = 0|h_1) = 1/5 = 0,2; P(A_3 = 1|h_1) = 4/5 = 0,8$
- $P(A_4 = 0|h_1) = 3/5 = 0,6; P(A_4 = 1|h_1) = 2/5 = 0,4$
- $P(A_1 = 0|h_2) = 0/5 = 0,0; P(A_1 = 1|h_2) = 5/5 = 1,0$
- $P(A_2 = 0|h_2) = 3/5 = 0,6; P(A_2 = 1|h_2) = 2/5 = 0,4$
- $P(A_3 = 0|h_2) = 4/5 = 0,8; P(A_3 = 1|h_2) = 1/5 = 0,2$
- $P(A_4 = 0|h_2) = 1/5 = 0,2; P(A_4 = 1|h_2) = 4/5 = 0,8$

Classificador Bayesiano Ingênuo

Exemplo

- Novo dado: $x = \langle 0, 0, 1, 1 \rangle$

Classificador Bayesiano Ingênuo

Exemplo

- Novo dado: $x = \langle 0, 0, 1, 1 \rangle$

$$h_{NB} = \operatorname{argmax}_i P(h_i) \prod_{j=1}^n P(a_j | h_i)$$

Classificador Bayesiano Ingênuo

Exemplo

- Novo dado: $x = \langle 0, 0, 1, 1 \rangle$

$$h_{NB} = \operatorname{argmax}_i P(h_i) \prod_{j=1}^n P(a_j | h_i)$$

$$P(H_{NB} = h_1) = P(h_1)P(a_1 = 0|h_1)P(a_2 = 0|h_1)P(a_3 = 1|h_1)P(a_4 = 1|h_1)$$

Classificador Bayesiano Ingênuo

Exemplo

- Novo dado: $x = \langle 0, 0, 1, 1 \rangle$

$$h_{NB} = \operatorname{argmax}_i P(h_i) \prod_{j=1}^n P(a_j | h_i)$$

$$P(H_{NB} = h_1) = P(h_1)P(a_1 = 0|h_1)P(a_2 = 0|h_1)P(a_3 = 1|h_1)P(a_4 = 1|h_1)$$

$$P(H_{NB} = h_1) = 0,5 \times 0,8 \times 0,8 \times 0,8 \times 0,4 = 0,1024$$

Classificador Bayesiano Ingênuo

Exemplo

- Novo dado: $x = \langle 0, 0, 1, 1 \rangle$

$$h_{NB} = \underset{i}{\operatorname{argmax}} P(h_i) \prod_{j=1}^n P(a_j | h_i)$$

$$P(H_{NB} = h_1) = P(h_1)P(a_1 = 0|h_1)P(a_2 = 0|h_1)P(a_3 = 1|h_1)P(a_4 = 1|h_1)$$

$$P(H_{NB} = h_1) = 0,5 \times 0,8 \times 0,8 \times 0,8 \times 0,4 = 0,1024$$

$$P(H_{NB} = h_2) = 0,5 \times 0,0 \times 0,6 \times 0,2 \times 0,8 = 0,0$$

Classificador Bayesiano Ingênuo

Exemplo

- Novo dado: $x = \langle 0, 0, 1, 1 \rangle$

$$h_{NB} = \underset{i}{\operatorname{argmax}} P(h_i) \prod_{j=1}^n P(a_j | h_i)$$

$$P(H_{NB} = h_1) = P(h_1)P(a_1 = 0|h_1)P(a_2 = 0|h_1)P(a_3 = 1|h_1)P(a_4 = 1|h_1)$$

$$P(H_{NB} = h_1) = 0,5 \times 0,8 \times 0,8 \times 0,8 \times 0,4 = 0,1024$$

$$P(H_{NB} = h_2) = 0,5 \times 0,0 \times 0,6 \times 0,2 \times 0,8 = 0,0$$

$$h_{NB} = \operatorname{argmax}(0,1024; 0,0) = 0,1024 = h_1 \Rightarrow h_1 \text{ é a escolhida}$$

Classificador Bayesiano Ingênuo

Exemplo – Classificação de Textos

Classificador Bayesiano Ingênuo

Exemplo – Classificação de Textos

- Tarefa:

Classificador Bayesiano Ingênuo

Exemplo – Classificação de Textos

- Tarefa:
 - A partir de exemplos de textos já classificados, classificar um novo texto

Classificador Bayesiano Ingênuo

Exemplo – Classificação de Textos

- Tarefa:
 - A partir de exemplos de textos já classificados, classificar um novo texto
 - Você tem o texto e não sabe a que classe pertence

Classificador Bayesiano Ingênuo

Exemplo – Classificação de Textos

- Tarefa:
 - A partir de exemplos de textos já classificados, classificar um novo texto
 - Você tem o texto e não sabe a que classe pertence
(retirou uma bala e não olhou, mas quer saber qual o sabor)

Classificador Bayesiano Ingênuo

Exemplo – Classificação de Textos

- Tarefa:
 - A partir de exemplos de textos já classificados, classificar um novo texto
 - Você tem o texto e não sabe a que classe pertence
(retirou uma bala e não olhou, mas quer saber qual o sabor)
- Função-alvo:

Classificador Bayesiano Ingênuo

Exemplo – Classificação de Textos

- Tarefa:
 - A partir de exemplos de textos já classificados, classificar um novo texto
 - Você tem o texto e não sabe a que classe pertence
(retirou uma bala e não olhou, mas quer saber qual o sabor)
- Função-alvo:
 - Definir se é spam (h_1) ou não (h_2)

Classificador Bayesiano Ingênuo

Exemplo – Classificação de Textos

- Tarefa:
 - A partir de exemplos de textos já classificados, classificar um novo texto
 - Você tem o texto e não sabe a que classe pertence
(retirou uma bala e não olhou, mas quer saber qual o sabor)
- Função-alvo:
 - Definir se é spam (h_1) ou não (h_2)
- Problemas:

Classificador Bayesiano Ingênuo

Exemplo – Classificação de Textos

- Tarefa:
 - A partir de exemplos de textos já classificados, classificar um novo texto
 - Você tem o texto e não sabe a que classe pertence
(retirou uma bala e não olhou, mas quer saber qual o sabor)
- Função-alvo:
 - Definir se é spam (h_1) ou não (h_2)
- Problemas:
 - Como representar um texto com atributos?

Classificador Bayesiano Ingênuo

Exemplo – Classificação de Textos

- Tarefa:
 - A partir de exemplos de textos já classificados, classificar um novo texto
 - Você tem o texto e não sabe a que classe pertence
(retirou uma bala e não olhou, mas quer saber qual o sabor)
- Função-alvo:
 - Definir se é spam (h_1) ou não (h_2)
- Problemas:
 - Como representar um texto com atributos?
 - Como estimar as probabilidades requeridas?

Classificador Bayesiano Ingênuo

Exemplo – Classificação de Textos

- Representação:

Classificador Bayesiano Ingênuo

Exemplo – Classificação de Textos

- Representação:
 - Um atributo A_i diferente para cada posição i de palavra no documento

Classificador Bayesiano Ingênuo

Exemplo – Classificação de Textos

- Representação:
 - Um atributo A_i diferente para cada posição i de palavra no documento
 - O valor do atributo ($A_i = a_i$) é a palavra nessa posição

Classificador Bayesiano Ingênuo

Exemplo – Classificação de Textos

- Representação:
 - Um atributo A_i diferente para cada posição i de palavra no documento
 - O valor do atributo ($A_i = a_i$) é a palavra nessa posição
- Exemplo:

Classificador Bayesiano Ingênuo

Exemplo – Classificação de Textos

- Representação:
 - Um atributo A_i diferente para cada posição i de palavra no documento
 - O valor do atributo ($A_i = a_i$) é a palavra nessa posição
- Exemplo:
 - “Aumente o tamanho do seu tênis em até 200cm!<p>Ligue agora para 0800-6969 e peça o moderníssimo PK aumentador Tabajara.”

Classificador Bayesiano Ingênuo

Exemplo – Classificação de Textos

- Representação:
 - Um atributo A_i diferente para cada posição i de palavra no documento
 - O valor do atributo ($A_i = a_i$) é a palavra nessa posição
- Exemplo:
 - “Aumente o tamanho do seu tênis em até 200cm!<p>Ligue agora para 0800-6969 e peça o moderníssimo PK aumentador Tabajara.”
 - 23 palavras \rightarrow 23 posições

Classificador Bayesiano Ingênuo

Exemplo – Classificação de Textos

- Representação:
 - Um atributo A_i diferente para cada posição i de palavra no documento
 - O valor do atributo ($A_i = a_i$) é a palavra nessa posição
- Exemplo:
 - “Aumente o tamanho do seu tênis em até 200cm!<p>Ligue agora para 0800-6969 e peça o moderníssimo PK aumentador Tabajara.”
 - 23 palavras \rightarrow 23 posições
 - O número de atributos aumenta com o número de palavras

Classificador Bayesiano Ingênuo

Exemplo – Classificação de Textos

- Assumamos que temos 700 spams e 300 não spams para treino

Classificador Bayesiano Ingênuo

Exemplo – Classificação de Textos

- Assumamos que temos 700 spams e 300 não spams para treino
- Recebemos um documento e, para classificá-lo:

Classificador Bayesiano Ingênuo

Exemplo – Classificação de Textos

- Assumamos que temos 700 spams e 300 não spams para treino
- Recebemos um documento e, para classificá-lo:

$$h_{NB} = \operatorname{argmax}_{h_i \in \{spam, \neg spam\}} P(h_i) \prod_{j=1}^{23} P(a_j | h_i)$$

Classificador Bayesiano Ingênuo

Exemplo – Classificação de Textos

- Assumamos que temos 700 spams e 300 não spams para treino
- Recebemos um documento e, para classificá-lo:

$$h_{NB} = \operatorname{argmax}_{h_i \in \{spam, \neg spam\}} P(h_i) \prod_{j=1}^{23} P(a_j | h_i)$$



$$h_{NB} = \operatorname{argmax}_{h_i \in \{spam, \neg spam\}} P(h_i) P(A_1 = \langle b \rangle | h_i) P(A_2 = \text{aumente} | h_i) \dots P(A_{23} = \text{tabajara} | h_i)$$

Classificador Bayesiano Ingênuo

Exemplo – Classificação de Textos

- Assumamos que temos 700 spams e 300 não spams para treino
- Recebemos um documento e, para classificá-lo:

$$h_{NB} = \operatorname{argmax}_{h_i \in \{spam, \neg spam\}} P(h_i) \prod_{j=1}^{23} P(a_j | h_i)$$



Cada $P(A_j = a_j | h_i)$
vem do treinamento

$$h_{NB} = \operatorname{argmax}_{h_i \in \{spam, \neg spam\}} P(h_i) P(A_1 = \langle b \rangle | h_i) P(A_2 = \text{aumente} | h_i) \dots P(A_{23} = \text{tabajara} | h_i)$$

Classificador Bayesiano Ingênuo

Exemplo – Classificação de Textos

- Maximizamos a probabilidade de observarmos as palavras do documento a ser classificado, na posição em que lá são encontradas

Classificador Bayesiano Ingênuo

Exemplo – Classificação de Textos

- Maximizamos a probabilidade de observarmos as palavras do documento a ser classificado, na posição em que lá são encontradas
- A hipótese (spam ou texto comum) que resultar em maior probabilidade será a vencedora

Classificador Bayesiano Ingênuo

Exemplo – Classificação de Textos

- Maximizamos a probabilidade de observarmos as palavras do documento a ser classificado, na posição em que lá são encontradas
- A hipótese (spam ou texto comum) que resultar em maior probabilidade será a vencedora
- Essa probabilidade, contudo, está sujeita à suposição de independência

Classificador Bayesiano Ingênuo

Exemplo – Classificação de Textos

- Independência:

Classificador Bayesiano Ingênuo

Exemplo – Classificação de Textos

- Independência:
 - As probabilidades de cada palavra em uma posição do texto são independentes das palavras que ocorrem em outras posições, dada a classificação do documento

Classificador Bayesiano Ingênuo

Exemplo – Classificação de Textos

- Independência:
 - As probabilidades de cada palavra em uma posição do texto são independentes das palavras que ocorrem em outras posições, dada a classificação do documento
- Falso!

Classificador Bayesiano Ingênuo

Exemplo – Classificação de Textos

- Independência:
 - As probabilidades de cada palavra em uma posição do texto são independentes das palavras que ocorrem em outras posições, dada a classificação do documento
- Falso!
 - A palavra “aumente”, se seguida de “seu tênis”, por exemplo, aumenta a probabilidade de ser spam

Classificador Bayesiano Ingênuo

Exemplo – Classificação de Textos

- Independência:
 - As probabilidades de cada palavra em uma posição do texto são independentes das palavras que ocorrem em outras posições, dada a classificação do documento
- Falso!
 - A palavra “aumente”, se seguida de “seu tênis”, por exemplo, aumenta a probabilidade de ser spam
- Contudo, se não assumirmos independência, poderemos ter que calcular um número abusivo de probabilidades

Classificador Bayesiano Ingênuo

Exemplo – Classificação de Textos

- Continuando o cálculo:

Classificador Bayesiano Ingênuo

Exemplo – Classificação de Textos

- Continuando o cálculo:
 - 1000 documentos, 700 spams, 300 não spams

Classificador Bayesiano Ingênuo

Exemplo – Classificação de Textos

- Continuando o cálculo:
 - 1000 documentos, 700 spams, 300 não spams
 - $P(spam) = 0,7$ e $p(\neg spam) = 0,3$

Classificador Bayesiano Ingênuo

Exemplo – Classificação de Textos

- Continuando o cálculo:
 - 1000 documentos, 700 spams, 300 não spams
 - $P(spam) = 0,7$ e $p(\neg spam) = 0,3$
 - $P(A_1 = \langle b \rangle | spam)$ é mais complicado

Classificador Bayesiano Ingênuo

Exemplo – Classificação de Textos

- Continuando o cálculo:
 - 1000 documentos, 700 spams, 300 não spams
 - $P(spam) = 0,7$ e $p(\neg spam) = 0,3$
 - $P(A_1 = \langle b \rangle | spam)$ é mais complicado
 - Podemos assumir então que a probabilidade de encontrar determinada palavra independe da posição

Classificador Bayesiano Ingênuo

Exemplo – Classificação de Textos

- Continuando o cálculo:
 - 1000 documentos, 700 spams, 300 não spams
 - $P(spam) = 0,7$ e $p(\neg spam) = 0,3$
 - $P(A_1 = \langle b \rangle | spam)$ é mais complicado
 - Podemos assumir então que a probabilidade de encontrar determinada palavra independe da posição

$$P(A_j = \omega_k | h_i) = P(\omega_k | h_i)$$

ω_k = cada palavra distinta no texto a ser classificado

Classificador Bayesiano Ingênuo

Exemplo – Classificação de Textos

- Continuando o cálculo:
 - 1000 documentos, 700 spams, 300 não spams
 - $P(spam) = 0,7$ e $p(\neg spam) = 0,3$
 - $P(A_1 = \langle b \rangle | spam)$ é mais complicado
 - Podemos assumir então que a probabilidade de encontrar determinada palavra independe da posição
$$P(A_j = \omega_k | h_i) = P(\omega_k | h_i)$$
$$\omega_k = \text{cada palavra distinta no texto a ser classificado}$$
 - O problema passa a ser calcular a probabilidade de achar determinada palavra, e não determinada palavra em certa posição

Classificador Bayesiano Ingênuo

Exemplo – Classificação de Textos

- Estimamos a probabilidade:

Classificador Bayesiano Ingênuo

Exemplo – Classificação de Textos

- Estimamos a probabilidade:

$$P(\omega_k|h_i) = \frac{n_k + 1}{n + |\text{vocabulário}|}$$

Classificador Bayesiano Ingênuo

Exemplo – Classificação de Textos

- Estimamos a probabilidade:

$$P(\omega_k|h_i) = \frac{n_k + 1}{n + |\text{vocabulário}|}$$

- $n \rightarrow$ número total de palavras (inclusive repetidas) em todos os exemplos de treino cuja classificação é h_i

Classificador Bayesiano Ingênuo

Exemplo – Classificação de Textos

- Estimamos a probabilidade:

$$P(\omega_k|h_i) = \frac{n_k + 1}{n + |\text{vocabulário}|}$$

- $n \rightarrow$ número total de palavras (inclusive repetidas) em todos os exemplos de treino cuja classificação é h_i
- $n_k \rightarrow$ número de vezes que a palavra ω_k é encontrada dentre essas n palavras

Classificador Bayesiano Ingênuo

Exemplo – Classificação de Textos

- Estimamos a probabilidade:

$$P(\omega_k|h_i) = \frac{n_k + 1}{n + |\text{vocabulário}|}$$

- $n \rightarrow$ número total de palavras (inclusive repetidas) em todos os exemplos de treino cuja classificação é h_i
- $n_k \rightarrow$ número de vezes que a palavra ω_k é encontrada dentre essas n palavras
- $|\text{vocabulário}| \rightarrow$ número total de palavras distintas (e outros tokens) encontrados nos dados de treinamento

Classificação de Textos

Qual a razão do |*vocabulário*|?

Classificação de Textos

Qual a razão do *vocabulário*?

- Suponha que o documento a ser classificado contenha uma palavra ω_k não existente nos exemplos de treino. Sem esse termo, a equação ficaria

$$P(\omega_k|h_i) = \frac{0}{n} = 0$$

Classificação de Textos

Qual a razão do *|vocabulário|*?

- Suponha que o documento a ser classificado contenha uma palavra ω_k não existente nos exemplos de treino. Sem esse termo, a equação ficaria

$$P(\omega_k|h_i) = \frac{0}{n} = 0$$

- **Suavização** (smoothing): forma de evitar essa radicalização

Classificação de Textos

Qual a razão do $|\text{vocabulário}|$?

- Suponha que o documento a ser classificado contenha uma palavra ω_k não existente nos exemplos de treino. Sem esse termo, a equação ficaria

$$P(\omega_k|h_i) = \frac{0}{n} = 0$$

- **Suavização** (smoothing): forma de evitar essa radicalização

$$P(\omega_k|h_i) = \frac{0 + 1}{n + |\text{vocabulário}|} = \frac{1}{n + |\text{vocabulário}|}$$

Classificação de Textos

Qual a razão do $|\text{vocabulário}|$?

- Suponha que o documento a ser classificado contenha uma palavra ω_k não existente nos exemplos de treino. Sem esse termo, a equação ficaria

$$P(\omega_k|h_i) = \frac{0}{n} = 0$$

- **Suavização** (smoothing): forma de evitar essa radicalização

$$P(\omega_k|h_i) = \frac{0 + 1}{n + |\text{vocabulário}|} = \frac{1}{n + |\text{vocabulário}|}$$

A probabilidade é baixíssima, mas existe!

Classificação de Textos

Suavização (Smoothing)

- Add-one smoothing (Laplace)

Classificação de Textos

Suavização (Smoothing)

- Add-one smoothing (Laplace)
 - A forma vista até agora

Classificação de Textos

Suavização (Smoothing)

- Add-one smoothing (Laplace)
 - A forma vista até agora

$$P(\omega_k|h_i) = \frac{n_k + 1}{\sum_k (1 + n_k)} = \frac{1 + n_k}{\sum_k n_k + \sum_k 1}$$

Classificação de Textos

Suavização (Smoothing)

- Add-one smoothing (Laplace)
 - A forma vista até agora

$$P(\omega_k|h_i) = \frac{n_k + 1}{\sum_k (1 + n_k)} = \frac{1 + n_k}{\sum_k n_k + \sum_k 1}$$

$$\Rightarrow P(\omega_k|h_i) = \frac{1 + n_k}{n + |\text{vocabulário}|}$$

Classificação de Textos

Suavização (Smoothing)

- Add-one smoothing (Laplace)
 - A forma vista até agora

$$P(\omega_k|h_i) = \frac{n_k + 1}{\sum_k (1 + n_k)} = \frac{1 + n_k}{\sum_k n_k + \sum_k 1}$$

$$\Rightarrow P(\omega_k|h_i) = \frac{1 + n_k}{n + |\text{vocabulário}|}$$

- Adicionamos 1 a cada palavra possível (vocabulário)

Suavização (Smoothing)

- Alternativa bastante usada
 - Adicionamos 1 ao numerador e 2 ao denominador

Suavização (Smoothing)

- Alternativa bastante usada
 - Adicionamos 1 ao numerador e 2 ao denominador

$$P(\omega_k | h_i) = \frac{n_k + 1}{n + 2}$$

Classificação de Textos

Suavização (Smoothing)

- Alternativa bastante usada
 - Adicionamos 1 ao numerador e 2 ao denominador

$$P(\omega_k|h_i) = \frac{n_k + 1}{n + 2}$$

A probabilidade de um termo desconhecido é $\frac{1}{(n + 2)}$

Classificador Bayesiano Ingênuo

Classificação de Textos – Algoritmo

Função *APRENDE*(*Exemplos*, $H = \{h_i\}$): **void**

Vocabulário \leftarrow conjunto de todas as diferentes palavras e tokens ocorridos em todo documento de *Exemplos*

para cada hipótese h_i **faça**

$docs_i \leftarrow$ subconjunto dos documentos de *Exemplos* para os quais a classificação é h_i

$$P(h_i) \leftarrow \frac{|docs_i|}{|Exemplos|}$$

$Texto_i \leftarrow$ um único documento criado pela concatenação de todos os membros de $docs_i$

$n \leftarrow$ número total de palavras em $Texto_i$

para cada palavra ω_k **no Vocabulário** **faça**

$n_k \leftarrow$ número de vezes que a palavra ω_k ocorre em $Texto_i$

$$P(\omega_k|h_i) \leftarrow \frac{(n_k + 1)}{(n + |Vocabulário|)}$$

Classificador Bayesiano Ingênuo

Classificação de Textos – Algoritmo

Função APRENDE(*Exemplos*, $H = \{h_i\}$): **void**

Vocabulário \leftarrow conjunto de todas as diferentes palavras e tokens ocorridos em todo documento de *Exemplos*

para cada hipótese h_i **faça**

$docs_i \leftarrow$ subconjunto dos documentos de *Exemplos* para os quais a classificação é h_i

$$P(h_i) \leftarrow \frac{|docs_i|}{|Exemplos|}$$

Conjunto de exemplos de treino

$Texto_i \leftarrow$ um único documento criado pela concatenação de todos os membros de $docs_i$

$n \leftarrow$ número total de palavras em $Texto_i$

para cada palavra ω_k **no Vocabulário** **faça**

$n_k \leftarrow$ número de vezes que a palavra ω_k ocorre em $Texto_i$

$$P(\omega_k|h_i) \leftarrow \frac{(n_k + 1)}{(n + |Vocabulário|)}$$

Classificador Bayesiano Ingênuo

Classificação de Textos – Algoritmo

Função *APRENDE*(*Exemplos*, $H = \{h_i\}$): **void**

Vocabulário \leftarrow conjunto de todas as diferentes palavras e tokens ocorridos em todo documento de *Exemplos*

para cada hipótese h_i **faça**

$docs_i \leftarrow$ subconjunto dos documentos de *Exemplos* para os quais a classificação é h_i

$$P(h_i) \leftarrow \frac{|docs_i|}{|Exemplos|}$$

Possíveis hipóteses

$Texto_i \leftarrow$ um único documento criado pela concatenação de todos os membros de $docs_i$

$n \leftarrow$ número total de palavras em $Texto_i$

para cada palavra ω_k **no Vocabulário** **faça**

$n_k \leftarrow$ número de vezes que a palavra ω_k ocorre em $Texto_i$

$$P(\omega_k|h_i) \leftarrow \frac{(n_k + 1)}{(n + |Vocabulário|)}$$

Classificador Bayesiano Ingênuo

Classificação de Textos – Algoritmo

Função $APRENDE(Exemplos, H = \{h_i\})$: **void**

Vocabulário \leftarrow conjunto de todas as diferentes palavras e tokens ocorridos em todo documento de *Exemplos*

para cada hipótese h_i **faça**

$docs_i \leftarrow$ subconjunto dos documentos de *Exemplos* para os quais a classificação é h_i

$$P(h_i) \leftarrow \frac{|docs_i|}{|Exemplos|}$$

Colete todas as palavras, pontuação e outros tokens que ocorrem nos exemplos

$Texto_i \leftarrow$ um único documento criado pela concatenação de todos os membros de $docs_i$

$n \leftarrow$ número total de palavras em $Texto_i$

para cada palavra ω_k **no Vocabulário** **faça**

$n_k \leftarrow$ número de vezes que a palavra ω_k ocorre em $Texto_i$

$$P(\omega_k|h_i) \leftarrow \frac{(n_k + 1)}{(n + |Vocabulário|)}$$

Classificador Bayesiano Ingênuo

Classificação de Textos – Algoritmo

Função *APRENDE*(*Exemplos*, $H = \{h_i\}$): **void**

Vocabulário \leftarrow conjunto de todas as diferentes palavras e tokens ocorridos em todo documento de *Exemplos*

para cada hipótese h_i **faça**

$docs_i \leftarrow$ subconjunto dos documentos de *Exemplos* para os quais a classificação é h_i

$$P(h_i) \leftarrow \frac{|docs_i|}{|Exemplos|}$$

← Calcule os termos de probabilidade $P(h_i)$ e $P(\omega_k|h_i)$

$Texto_i \leftarrow$ um único documento criado pela concatenação de todos os membros de $docs_i$

$n \leftarrow$ número total de palavras em $Texto_i$

para cada palavra ω_k **no Vocabulário** **faça**

$n_k \leftarrow$ número de vezes que a palavra ω_k ocorre em $Texto_i$

$$P(\omega_k|h_i) \leftarrow \frac{(n_k + 1)}{(n + |Vocabulário|)}$$

Classificador Bayesiano Ingênuo

Classificação de Textos – Algoritmo

Função *CLASSIFICA*(*Doc*, *Vocabulário*): a hipótese estimada para o documento *Doc*

Palavras \leftarrow todas as palavras em *Doc* que também são encontradas no *Vocabulário*

$$h_{NB} = \operatorname{argmax}_{h_i \in H} P(h_i) \prod_{a_j \in \text{Palavras}} P(a_j | h_i)$$

retorna h_{NB}

Classificador Bayesiano Ingênuo

Classificação de Textos – Algoritmo

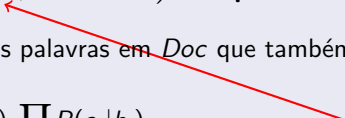
Função *CLASSIFICA*(*Doc*, *Vocabulário*): a hipótese estimada para o documento *Doc*

Palavras \leftarrow todas as palavras em *Doc* que também são encontradas no *Vocabulário*

$$h_{NB} = \operatorname{argmax}_{h_i \in H} P(h_i) \prod_{a_j \in \text{Palavras}} P(a_j | h_i)$$

retorna h_{NB}

Documento a ser classificado



Classificador Bayesiano Ingênuo

Classificação de Textos – Algoritmo

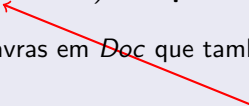
Função *CLASSIFICA*(*Doc*, *Vocabulário*): a hipótese estimada para o documento *Doc*

Palavras \leftarrow todas as palavras em *Doc* que também são encontradas no *Vocabulário*

$$h_{NB} = \operatorname{argmax}_{h_i \in H} P(h_i) \prod_{a_j \in \text{Palavras}} P(a_j | h_i)$$

retorna h_{NB}

Palavras distintas encontradas durante o treino



Classificador Bayesiano Ingênuo

Classificação de Textos – Algoritmo


Função *CLASSIFICA*(*Doc*, *Vocabulário*): a hipótese estimada para o documento *Doc*

Palavras \leftarrow todas as palavras em *Doc* que também são encontradas no *Vocabulário*

$$h_{NB} = \operatorname{argmax}_{h_i \in H} P(h_i) \prod_{a_j \in \text{Palavras}} P(a_j | h_i)$$

retorna h_{NB}

Palavras novas são ignoradas



Classificador Bayesiano Ingênuo

Classificação de Textos – Algoritmo

Função *CLASSIFICA*(*Doc*, *Vocabulário*): a hipótese estimada para o documento *Doc*

Palavras \leftarrow todas as palavras em *Doc* que também são encontradas no *Vocabulário*

$$h_{NB} = \operatorname{argmax}_{h_i \in H} P(h_i) \prod_{a_j \in \text{Palavras}} P(a_j | h_i) \leftarrow a_i \text{ denota } i\text{-ésima palavra em } Doc$$

retorna h_{NB}

Classificador Bayesiano Ingênuo

Classificação de Textos – Modificações

- Podemos ignorar as palavras mais comuns (um, a, o, de etc)

Classificador Bayesiano Ingênuo

Classificação de Textos – Modificações

- Podemos ignorar as palavras mais comuns (um, a, o, de etc)
- Podemos retirar do vocabulário qualquer palavra que tenha ocorrido menos vezes que um mínimo (menos que 2 vezes, por exemplo)

Referências

- Russell, S.; Norvig P. (2010): Artificial Intelligence: A Modern Approach. Prentice Hall. 3a ed.
 - Slides do livro: aima.eecs.berkeley.edu/slides-pdf/
- Mitchell, T.M.: Machine Learning. McGraw-Hill. 1997.
- Murphy, K. P.: Machine Learning: A Probabilistic Perspective. MIT Press. 2012.
- Alpaydm, E.: Introduction to Machine Learning. 2 ed. MIT Press. 2010.