**I. Pen-and-paper**

1. Vamos calcular uma época do algoritmo Expectation Maximization Clustering, de forma a atualizar os clusters fornecidos no enunciado, para que melhor classifiquem o dataset. Para isso vamos aplicar os seguintes dois passos:

1º - Atribuir classes aos pontos

Usando o teorema de Bayse vamos calcular a probabilidade de cada ponto pertencer a um dos clusters

Antes de calcular o posteior temos de determinar as seguintes parcelas:

- Likelihoods = d

- Joint Probability =

Todos os parâmetros necessários, são fornecidos no enunciado falta então apenas calcular os posteriors para cada instância do dataset, estas probabilidades é que vão definir a qual cluster a instância pertence

-

*Cluster*

- **Likelihood**:

- **Joint Probability**:

*Cluster*

- **Likelihood**:

- **Joint Probability**:

- **Denominador**:

- **Posterior**: - **Posterior**:

-

*Cluster*

- **Likelihood**:

- **Joint Probability**:

*Cluster*

- **Likelihood**:

- **Joint Probability**:

- **Denominador**:

- **Posterior**: - **Posterior**:

-

*Cluster*

- **Likelihood**:

- **Joint Probability**:

*Cluster*

- **Likelihood**:

- **Joint Probability**:

- **Denominador**:

- **Posterior**: - **Posterior**:

-

*Cluster*

- **Likelihood**:

- **Joint Probability**:

*Cluster*

- **Likelihood**:

- **Joint Probability**:

- **Denominador**:

- **Posterior**: - **Posterior**:

|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  |  |
|  |  |  |  |

Analisando o posterior calculado, podemos verificar que as instâncias do dataset pertencem aos seguintes clusters:

2º - Maximization step (atualizar os parâmetros que definem o cluster de forma a melhor separar o dataset)

De forma a tornar os cálculos mais simples, vamos organizar os posteiros em dois vetores (um para cada cluster)

Estimar Priors

Estimar

Estimar

A picture containing diagram

Description automatically generated

Esta imagem mostra a forma dos novos clusters calculados usando o método EM

1. Usando as classificações dos pontos calculadas no exercício anterior podemos então calcular a silhueta de cada cluster e consequentemente a geral

Onde:

Distância média entre x e os pontos que pertencem ao mesmo cluster

Distância média entre x e os pontos do cluster mais próximo

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  |  |  |  |
|  | 4.038843615 | 0 (único ponto do cluster) | 4.495358041 | 4.928650381 |
|  | 8.544003745 | 6.982375994 | 6 | 6.403124237 |
|  | 0.52728911 | 1 | 0.25077366 | 0.230274129 |

A silhueta de um cluster e dada pela média aritmética das silhuetas dos pontos que lhe pertencem, assim:

Da mesma forma, a silhueta do conjunto de clusters é a média aritmética da silhueta de cada cluster

1. De forma simplificar a resolução deste problema vamos primeiro aproximar a VC-dimension para um espaço geral com N features. Esta aproximação vai ser feita usando o número de parâmetros necessários para cada classificador, já que:

i.) MLP com 3 hidden layers onde cada camada possui N (número de features) nodes

1. Temos uma matriz , parâmetros, e um vetor parâmetros
2. Temos uma matriz , parâmetros, e um vetor parâmetros
3. Temos uma matriz , parâmetros, e um vetor parâmetros
4. Temos uma matriz parâmetros, e um vetor parâmetros

Assim, a classificação usando uma MLP necessita de parâmetros que aproxima o valor da VC-dimension

ii) Árvore de Decisão onde as variáveis são discreteadas em 3 bins

Assim o número total de combinações possíveis (3 possibilidades para cada feature). Para classificar estas instâncias basta contruir onde cada folha contém a classificação para uma destas instâncias. Devido a natureza da árvore de decisão conseguimos classificar perfeitamente qualquer um destes pontos (cada

um vai corresponder a um ramo da árvore), assim podemos concluir que a VC-dimension da árvore de decisão é igual a .

iii) Baysian classifier com likelihoods com distribuição normal multivariada

Como:

Assim os parametros das likelihoods são os mesmo da normal multivariada:

- = vetor de médias (), corresponde então a N parâmetros distintos

- = matriz de covariância (), mas a matriz é simétrica logo apenas corresponde a parâmetros distintos

Assim uma classificação Baysiana necessita de parâmetros que a próxima o valor da VC-dimension

Tendo então obtido uma fórmula para a VC-dimension em função no número de features torna a resolução dos seguintes exercícios trivial

1. Para N = 5, temos que
2. VC-dimension = 102
3. VC-dimension = 243
4. VC-dimension = 41
5. Chart, line chart

   Description automatically generatedChart, line chart

   Description automatically generatedPodemos verificar que a VC-dimension da árvore de decisão é exponencialmente maior que os restantes dois classificadores
6. O grau de crescimento dos dois classificadores é semelhante (crescimento quadrático)

1. Answer 4

**II. Programming and critical analysis**

1. Answer 5
2. Answer 6
3. Answer 7
4. Answer 8

**III. APPENDIX**

Paste your programming code here using Consolas 9pt or 10pt.

Use **highlighting** or colored text to facilitate the analysis by your faculty hosts.

**END**