ECE421 Lab3 Report

1.K-means

1.1 Learning K-means

1)

distance\_func

图片包含 日程表

描述已自动生成

Implement K-means

散点图

中度可信度描述已自动生成

文本

描述已自动生成

文本

描述已自动生成

图表

描述已自动生成

2)

文本

描述已自动生成

图片包含 Word

描述已自动生成

Plot

图表, 散点图

描述已自动生成

图表, 散点图

描述已自动生成

图表, 散点图

描述已自动生成

图表, 散点图, 气泡图

描述已自动生成

图表, 散点图, 气泡图

描述已自动生成

From the figures above,

|  |  |
| --- | --- |
| Number of Cluster - K | Validation Loss |
| 1 | 3.861 |
| 2 | 0.888 |
| 3 | 0.488 |
| 4 | 0.316 |
| 5 | 0.275 |

The minimum validation loss achieved here is when K=5, validation loss = 0.275

The validation loss decreases as K increase.

However, it doesn’t make sense to divide the dataset into too many clusters.

The validation loss decreases drastically as K increase from 1 to 3, the validation loss drops 3.4.

The validation loss decreases much slower when increase K from 3 to 5, the validation loss drops 2.1.

Thus, I think K =3 is the optimal number of clusters.

2.Mixtures of Gaussians

2.1 Gaussian cluster mode

1)

图片包含 文本

描述已自动生成

2)

图片包含 文本

描述已自动生成

The reason to use reduce\_logsumexp since we need to compute

while tf. reduce\_sum only compute the sum

2.2 Learning the MoG

1)

散点图

描述已自动生成

文本

描述已自动生成

图形用户界面, 文本

描述已自动生成

Plot Loss:

图表, 折线图

描述已自动生成

Model Parameter:

|  |  |  |  |
| --- | --- | --- | --- |
| Cluster K | Log pi | mu | Sigma |
| 1 | -1.0067024 | [ 0.05538136 -1.6027924] | 2.9715424 |
| 2 | -1.1738701 | [-1.1191827 -3.316661] | 0.07645379 |
| 3 | -1.1226696 | [ 1.3025047 0.31507367] | 0.09755903 |

2)

图片包含 散点图

描述已自动生成

图表, 散点图

描述已自动生成

图表, 散点图

描述已自动生成

图表, 散点图

描述已自动生成

图表, 散点图, 气泡图

描述已自动生成

图表, 散点图, 气泡图

描述已自动生成

From the figures above,

|  |  |
| --- | --- |
| Number of Cluster - K | Validation Loss |
| 1 | 2.87 |
| 2 | 2.25 |
| 3 | 1.945 |
| 4 | 1.458 |
| 5 | 1.454 |

The minimum validation loss achieved here is when K=5, validation loss = 1.454

The validation loss decreases as K increase.

However, it doesn’t make sense to divide the dataset into too many clusters.

The validation loss decreases drastically as K increase from 1 to 3, the validation loss drops 0.93.

The validation loss decreases much slower when increase K from 3 to 5, the validation loss drops 0.39.

Thus, I think K =3 is the optimal number of clusters.

3) Compare K-means and MoG

|  |  |  |
| --- | --- | --- |
| K | k-means Validation Loss | MoG Validation Loss |
| 5 | 21.54334444037717 | 12.697508813381338 |
| 10 | 21.24050911505024 | 10.834202951545155 |
| 15 | 21.042035154237134 | 9.654315040879087 |
| 20 | 20.992872165406673 | 8.715772553817882 |
| 30 | 20.71745446427862 | 6.316014218609361 |

The minimum validation loss for both k-means and MoG achieved here is when K=30.

k-means validation loss = 20.717 MoG validation loss = 6.316

The validation loss for both k-means and MoG decreases as K increase.

However, increase in K post more effect on MoG algorithm, since its validation loss decrease more rapidly comparing to the validation loss for k-means as K increase.

From the graph above , we can see that both k-means and Mog validation loss decrease with a steady rate as K increase thus, I think K =30 is the optimal number of clusters.