資料科學導論 exetcise4 大氣4A 黃展皇 106601015

程式碼如附件code\_106601015.py

Requestment：matplotlib、numpy、pandas

學生作業實作的部分是實現MF\_template.py中mf\_sgd方法對每個epoch的每個i, j, r疊代的code內容，如下：

# Predict current rating

            pred = b + b\_u[i] + b\_i[j] # r\_hat\_ij = 𝜇+b\_i+c\_j + p\_i\*q\_j

            for k in range(K):

                pred += P[i, k] \* Q[j, k]

            err = r - pred # 𝑑\_𝑖𝑗=r\_ij−r\_hat\_ij

            # Update biases

            b\_u[i] += alpha \* (err - beta \* b\_u[i]) # −𝜂\*∇(𝑏\_𝑖)=𝜂(𝑑\_𝑖𝑗-𝜆𝑏\_𝑖)

            b\_i[j] += alpha \* (err - beta \* b\_i[j])

            # Update latent factors

            for k in range(K):

                Pf = P[i, k]

                Qf = Q[j, k]

                P[i, k] += alpha \* (err \* Qf - beta \* Pf) # −𝜂\*∇(P)=𝜂(𝑑\_𝑖𝑗q\_j-𝜆p\_i)

                Q[j, k] += alpha \* (err \* Pf - beta \* Qf)

1. 程式碼講解與實作執行：

實作的部分有兩大目標：

1. update "b\_u"(vector of rating bias for users) and "b\_i"(vector of rating bias for items)
2. update user and item latent feature matrices "P", "Q"

程式第一段用b + b\_u[i] + b\_i[j] + sum(P[i, k] \* Q[j, k])產生預測值pred，並計算r – pred為err。

程式第二段用梯度下降loss算法計算並更新biases(b\_u, b\_i)->完成目標(1)

程式第三段用用梯度下降loss算法計算並更新對所有k的latent factors(P, Q) ->完成目標(2)

執行結果：

Iteration: 10 /20/30/40/50; error = 0.9782/0.9526/0.9369/0.9258/0.9174

P x Q:<秀出訓練後的pred>

Global bias:3.501556983616962

User bias: <秀出訓練後的b\_u>

Item bias: <秀出訓練後的b\_i>

1. RMSE error值結果討論：

疊代第一次時RMSE error大約為1.03，而後隨著疊代次數增加，RMSE error也相應減少，不過d(RMSE error)/d(item)由負值逐步趨於0，在item=50次時RMSE error大約下降到0.91，如圖：

