Q1 (3 pts): (A) B C (D) **Q2 (3 pts):** (A) B C D (E)

Q3 (5 pts): (A) B (C) (D) E **Q4 (10 pts):** A (B) (C) D (E)

Q5 (6 pts):

Score = $(0.5) + (-0.7) + (-0.2) + (-0.1)$

$= (-0.5)$

Prediction = sign(score) = sign$(-0.5) = -1$

Q6 (15 pts): a) (i) $\frac{\lambda_1}{\lambda_1+\lambda_2+\lambda_3} \times 100\%$ (ii) $\frac{\lambda_1+\lambda_2}{\lambda_1+\lambda_2+\lambda_3} \times 100\%$ (iii) 100%

b) $\hat{x} = \vec{\mu} + ((\vec{x}-\vec{\mu})\cdot\vec{v}_1)\vec{v}_1 + ((\vec{x}-\vec{\mu})\cdot\vec{v}_2)\vec{v}_2$

c) $\vec{u} = \vec{v}_1$ [Eigen vector with highest λ]

Q7 (8 pts):

$\nabla F(\vec{x})\big|_{(0,0,0,0)} = (4, -1, -256, 1)$

Calculations here:

$\left(\frac{\partial F}{\partial x_1}\right)_{x_1=0} = 2(x_1+2) = 4$

$\left(\frac{\partial F}{\partial x_2}\right)_{x_2=0} = \frac{\partial}{\partial x_2}(1-x_2) = -1$

$\left(\frac{\partial F}{\partial x_3}\right)_{x_3=0} = 4(x_3-4)^3 = 4(-4)^3 = -256$

$\left(\frac{\partial F}{\partial x_4}\right)_{x_4=0} = \frac{\partial}{\partial x_4}(x_4+1) = 1$

Q8 (10 pts):

Method 1 is faster as it computes mean & variance in 1 pass over RDD.

Method 2 makes 2 passes over RDD, first to calculate Mean, second to calculate Var. hence is slower.

Q9 (14 pts):

```
N, D = graphRDD.groupByKey().
                .mapValues(lambda x: len(x))
                .sortByKey(ascending=False)
                .first()
                .sortBy(lambda x: x[1], False)
                .first()
```

Q10 (6 pts):

```
customersRDD.fullOuterJoin(purchaseRDD).collect()
```

OR.

```
customerRDD.leftOuterJoin(purchaseRDD.).collect()
```