Quiz 8

1)

1. $w_{2,3} = \frac{(4-2.5)(3-2.5)+(1-2.5)(2-2.5)}{\sqrt{(4-2.5)^2+(1-2.5)^2}\times\sqrt{(3-2.5)^2+(2-2.5)^2}} = 1$

$w_{3,4} = 0$

$$w_{3,5} = \frac{(3-3)\left(2-\frac{10}{3}\right)+(2-3)\left(3-\frac{10}{3}\right)+(4-3)(5-\frac{10}{3})}{\sqrt{(3-3)^2+(2-3)^2+(4-3)^2}\times\sqrt{(2-\frac{10}{3})^2+(3-\frac{10}{3})^2+(5-\frac{10}{3})^2}} = 0.65$$

$P_{2,3} = \frac{\left(2-\frac{5}{2}\right)\times 1+(4-4)\times 0+(1-\frac{10}{3})\times 0.65}{1+0+0.65} + 3 = 1.78$ (1pt, no need to show the result)

2. No, because U1 has no rating on I2 (1pt)

3. Jaccard Sim(U1, U3) = 1 (1pt)

2) 1.

$$w_{1,3} = \frac{\left(3-\frac{10}{3}\right)\left(5-\frac{8}{3}\right)+(5-\frac{10}{3})\left(1-\frac{8}{3}\right)}{\sqrt{(3-\frac{10}{3})^2+(5-\frac{10}{3})^2}\times\sqrt{(5-\frac{8}{3})^2+(1-\frac{8}{3})^2}} = -0.73$$

$$w_{2,3} = \frac{\left(4-\frac{8}{3}\right)\left(2-\frac{8}{3}\right)+(3-\frac{8}{3})\left(1-\frac{8}{3}\right)}{\sqrt{(4-\frac{8}{3})^2+(3-\frac{8}{3})^2}\times\sqrt{(2-\frac{8}{3})^2+(1-\frac{8}{3})^2}} = -0.58$$

$$w_{3,4} = \frac{\left(5-\frac{8}{3}\right)\left(2-\frac{8}{3}\right)+(2-\frac{8}{3})\left(3-\frac{8}{3}\right)}{\sqrt{(5-\frac{8}{3})^2+(2-\frac{8}{3})^2}\times\sqrt{(2-\frac{8}{3})^2+(3-\frac{8}{3})^2}} = -0.98$$

(0.5pt, no need to show the result to get full credit)

$P_{1,3} = \frac{2\times w_{1,3}+1\times w_{2,3}}{|w_{1,3}|+|w_{2,3}|}$ (0.5pt, no need to show the result.)

3) Advantage: Reduce the computation (1pt)

Disadvantage: May lose useful information (1pt)

4) Inverse frequency. $f_i = \log(N/n_i)$ (1pt)

e.g) clustering --- get rid of the outsiders. Case Amplification (1pt)

5) Advantage: Simple representation (1pt)

Disadvantage: Complicated for large dataset (1pt)