CSE881 HW1

Nan Cao, A52871775

Sep 9th, 2016

Problem 1

(a): Discrete; Qualitative; Nominal;
(b): Discrete; Qualitative; Ordinal;
(c): Continuous; Quantitative; Ratio;
(d): Continuous; Quantitative; Interval;
(e): Continuous; Quantitative; Interval;
(f): Discrete; Qualitative; Ordinal

Problem 2

(a): Yes, because salary of computer engineers is an ordinal attribute, and median works good on it.
(b): Yes, because weight and height of individuals are ratio attribute, and correlation is proper to ratio attributes.
(c): No, because Richter Magnitude Scale is an ordinal attribute, and it's a logarithmic scale instead of a linear one.
(d): Yes, because gender is a nominal attribute, and entropy is proper to this kind of data.
(e): No, because degree temperature in Degree Celsius is an interval attribute, geometric doesn't work on it.
(f): No, because GPA is an ordinal attribute, standard deviation doesn't work on it.

Problem 3

(You can find my source code and outputs at the end of the homework.)

$$y_x = (0.2901, 0.2574, 0.7762)$$
$$w = (-0.7960, 0.5970, 0.0995)$$

Problem 4

(a):

$$\rho = \begin{bmatrix} 1 & \frac{199.37}{\sqrt{389.75*610.52}} & \frac{135.12}{\sqrt{389.75*359.36}} \\ \frac{199.37}{\sqrt{389.75*610.52}} & 1 & \frac{426.30}{610.52*359.36} \\ \frac{135.12}{\sqrt{389.75*359.36}} & \frac{426.30}{610.52*359.36} & 1 \end{bmatrix}$$
$$= \begin{bmatrix} 1 & 0.4087 & 0.3610 \\ 0.4087 & 1 & 09101 \\ 0.3610 & 0.9101 & 1 \end{bmatrix}$$

Yes, according to the correlation matrix, weight is more correlated to age.

(b):

No, the covariance will remain the same.

$$\begin{aligned} A &= age;\ W = weight \\ W' &= centered\ weight = W - \bar{W} = W - E(W); \\ E(W') &= E(W - \bar{W}) = 0 \\ Cov(A, W') &= E(A - E(A))E(W' - E(W')) \\ &= E(A - E(A))E(W' - 0) \\ &= E(A - E(A))E(W - E(W)) \\ &= Cov(A, W) \end{aligned}$$

(c):

Covariance between weight (in kilogram) and age be smaller than 199.37.

$$\begin{aligned} A &= age \\ W_p &= weight\ (in\ pounds) \\ W_k &= weight\ (in\ kilograms) \\ W_p &= 2.2W_k;\ W_k = \frac{5}{11}W_p \\ Cov(A, W_k) &= Cov(A, \frac{5}{11}W_p) \\ &= E(A - E(A))E(\frac{5}{11}W_p) - E(\frac{5}{11}W_p))) \\ &= \frac{5}{11}E(A - E(A))E(W_p - E(W_p)) \\ &= \frac{5}{11}Cov(A, W_p) \\ &= \frac{5}{11} * 199.37 \\ &= 54.2591 < 119.37 \end{aligned}$$

(d):
Their covariance value be 199.37

$$
\begin{aligned}
A &= age;\ W = weight \\
A^* &= \frac{A-\bar{A}}{\sigma_A} \\
W^* &= \frac{W-\bar{W}}{\sigma_W} \\
Cov(A^*,W^*) &= E(A^*-E(A^*))E(W^*-E(W^*)) \\
&= E(\frac{A-\bar{A}}{\sigma_A} - E(\frac{A-\bar{A}}{\sigma_A}))E(\frac{W-\bar{W}}{\sigma_W} - E(\frac{W-\bar{W}}{\sigma_W})) \\
&= \frac{E(A-E(A))E(W-E(W))}{\sigma_A\sigma_W} \\
&= \frac{Cov(A,W))}{\sigma_A\sigma_W} = \frac{199.37}{389.75 * 610.52} < 199.37
\end{aligned}
$$

Problem 5

$$
\begin{aligned}
p &= \frac{2NQ + Z^2_{\frac{\alpha}{2}} \pm Z_{\frac{\alpha}{2}}\sqrt{Z^2_{\frac{\alpha}{2}} + 4NQ - 4NQ^2}}{2(N + Z^2_{\frac{\alpha}{2}})} \\
&= \frac{2*150*0.95 + 1.96^2 \pm 1.96\sqrt{1.96^2 + 4*150*0.95 - 4*150*1.96^2}}{2(150+1.96^2)} \\
&= (0.90254, 0.97499)
\end{aligned}
$$

Yes, it is safe to conclude (say, with 95% confidence) that my method outperforms the baseline method.

Source Codes & Output for Problem 3

Source Codes

```
X=[0.1,0.1,0.2;0.5,0.6,0.4]'
y=[0.1,0.4,0.8]'
 [U S V]=svd(X)
A=U(1:3,1:2)
P=A*(A'*A)^(-1)*A'
yx=P*y
w=cross(X(1:3,1),X(1:3,2))
w=w/norm(w)
```

Outputs

```
X =

    0.1000    0.5000
    0.1000    0.6000
```

```
    0.2000    0.4000


y =

    0.1000
    0.4000
    0.8000


U =

   -0.5632   -0.2216   -0.7960
   -0.6705   -0.4404    0.5970
   -0.4829    0.8700    0.0995


S =

    0.9042         0
         0    0.1111
         0         0


V =

   -0.2432    0.9700
   -0.9700   -0.2432


A =

   -0.5632   -0.2216
   -0.6705   -0.4404
   -0.4829    0.8700


P =

    0.3663    0.4752    0.0792
    0.4752    0.6436   -0.0594
    0.0792   -0.0594    0.9901


yx =

    0.2901
    0.2574
    0.7762


w =

   -0.0800
```

```
    0.0600
    0.0100


w =

   -0.7960
    0.5970
    0.0995
```