1-

الف)

وقتی $k = 3$ باشد کمترین میزان خطا را داریم زیرا سیگمای خطا داده های اموزش حداقل می شود

ب)

بهنیه ترین $k = n$ ^1/2 است که n تعداد نمونه های ماست و k بهتر است فرد باشد و ضریبی از تعداد کلاس های موجود نباشد، مثلا در این مثال $k = \sqrt{10} = 3.16227$ است

2-هردو discriminative هستند.در DT مرز بین کلاس های پیدا شده و براسا ان برای داده های جدید دسته بندی انجام می شود، و KNN با توجه به احتمال شرطی دسته بندی داده های جدید را انجام میدهد و هیچکدام به توزیع داده های برای دسته بندی کاری ندارند.

۲- الف)

$$\frac{d}{dx}\sigma(x) = \frac{d}{dx}\frac{1}{1+e^{-x}} = \frac{d}{dx}(1+e^{-x})^{-1} = -(1+e^{-x})\frac{d}{dx}(1+e^{-x}) =$$

$$-(1+e^{-x})^{-2}\cdot\frac{d}{dx}[e^{-x}] = -(1+e^{-x})^{-2}\cdot\left(e^{-x}\cdot\frac{d}{dx}-x\right) = -(1+e^{-x})^{-2}\cdot$$

$$(e^{-x}\cdot -1) = (1+e^{-x})^{-2}e^{-x} = \frac{e^{-x}}{(1+e^{-x})(1+e^{-x})} = \frac{1}{1+e^{-x}}\cdot$$

$$\left(\frac{1+e^{-x}}{1+e^{-x}} - \frac{1}{1+e^{-x}}\right) = \frac{1}{1+e^{-x}}\cdot\left(1-\frac{1}{1+e^{-x}}\right) = \underline{\sigma(x)\cdot(1-\sigma(x))}$$

ب)

$$P(y^i=0|x^i) = 1-\frac{1}{1+e^{-w^Tx^i}}, \quad P(y^i=1|x^i) = \frac{1}{1+e^{-w^Tx^i}}$$

$$P(y^i|x^i) = \left(\frac{1}{1+e^{-w^Tx^i}}\right)^{y^i}\left(\frac{e^{-w^Tx^i}}{1+e^{-w^Tx^i}}\right)^{1-y^i} \qquad y^i\in\{0,1\}$$

$$P(y|x^i) = \prod_{i=1}^{n}\left(\frac{1}{1+e^{-w^Tx^i}}\right)^{y^i}\left(\frac{e^{-w^Tx^i}}{1+e^{-w^Tx^i}}\right)^{1-y^i}$$

$$\xrightarrow{\log} -\lg P(y|x^i) = -\lg\prod_{i=1}^{n}\left(\frac{1}{1+e^{-w^Tx^i}}\right)^{y^i}\left(\frac{e^{-w^Tx^i}}{1+e^{-w^Tx^i}}\right)^{1-y^i} \rightarrow$$

$$\underline{-\left(\sum_{i=1}^{n} y^i\lg\left(\frac{1}{1+e^{-w^Tx^i}}\right) + (1-y^i)\lg\left(\frac{e^{-w^Tx^i}}{1+e^{-w^Tx^i}}\right)\right)}$$

ج)

$$-\left[\sum_{i=1}^{n} y^i \lg\frac{1}{1+e^{-w^Tx^i}} + (1-y^i)\lg\frac{1}{1+e^{-w^Tx^i}}\right] \longrightarrow$$

$$\left[\sum_{i=1}^{n} y^i x^i \frac{e^{-w^Tx^i}}{1+e^{-w^Tx^i}} - (1-y^i)x^i\frac{e^{-w^Tx^i}}{1+e^{-w^Tx^i}}\right] = \sum_{i=1}^{n} x^i\Big[$$

$$\frac{y^i - e^{-w^Tx^i}}{1+e^{-w^Tx^i}} - \frac{1}{1+e^{-w^Tx^i}} + \frac{y^i}{1+e^{-w^Tx^i}}\Big] = \sum\Big[y^i\Big[\frac{e^{-w^Tx^i}+1}{e^{-w^Tx^i}+1}\Big] -$$

$$\frac{1}{1+e^{-w^Tx^i}}\Big]x^i = \sum\left[y^i - \underline{\sigma(w^Tx^i)}\right]x^i$$

د)اگر منظور از بیش براز در این مثال است، با افزایش ضرایب(w^i ها) مدل ما مستعد بیش برازش می شود چون پیچیدگی مدل افزایش می یابد.

(۴_الف)

X_1:

$P(yes) = \frac{9}{14}$ $P(NO) = \frac{5}{14}$

$P(youth \mid Y) = \frac{2}{9}$

$P(high \mid Y) = \frac{2}{9}$

$P(yes \mid Y) = \frac{6}{9}$

$P(Fair \mid Y) = \frac{6}{9}$

0.0001

$P(Youth \mid N) = \frac{3}{5}$

$P(high \mid N) = \frac{2}{5}$

$P(yes \mid NO) = \frac{1}{5}$

$P(Fair \mid N) = \frac{2}{5}$

0.006

$\boxed{NO}$

X_2:

$P(Senior \mid Y) = \frac{3}{9}$

$P(Low \mid Y) = \frac{3}{9}$

$P(NO \mid Y) = \frac{3}{9}$

$P(excellent \mid Y) = \frac{3}{9}$

0.007

$P(Senior \mid N) = \frac{2}{5}$

$P(Low \mid N) = \frac{1}{5}$

$P(NO \mid N) = \frac{4}{5}$

$P(excellent \mid N) = \frac{3}{5}$

0.013

$\boxed{NO}$

X_3:

$P(Middle\ age \mid Y) =$

$P(Medium \mid Y) =$

$P(No \mid Y) =$

$P(Fair \mid Y) =$

$P(Middle\ age \mid N) = 0$

$P(medium \mid N) = \frac{2}{5}$

$P(NO \mid N) = \frac{4}{5}$

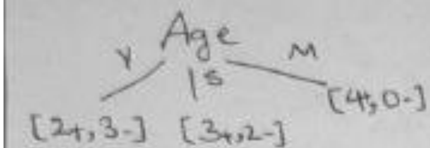
$P(Fair \mid N) = \frac{2}{5}$

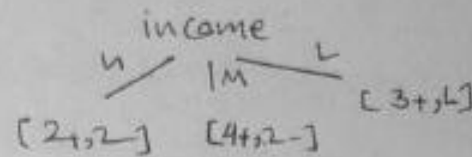
$\boxed{Yes}$

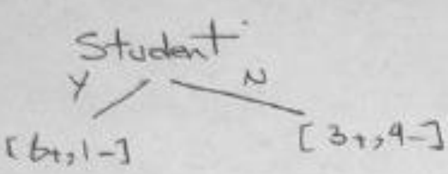
$S = [9+, 5-]$

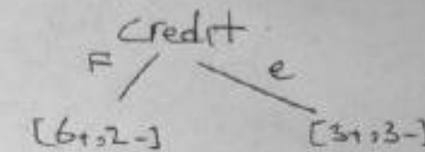
(۴-ب

$$\text{Entropy}(S) = -\left(\frac{9}{14}\log_2 \frac{9}{14} + \frac{5}{14}\log_2 \frac{5}{14}\right) = 0.94$$









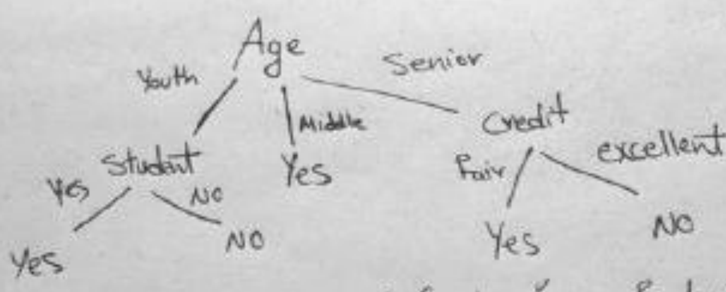
$$\text{Gain}(S, Age) = 0.94 - \frac{5}{14}\left(\frac{2}{5}\log_2\frac{2}{5} + \frac{3}{5}\log_2\frac{3}{5}\right) - \frac{5}{14}\left(\frac{3}{5}\lg_2\frac{3}{5} + \frac{2}{5}\lg_2\frac{2}{5}\right) - \frac{4}{14}\left(\frac{4}{4}\lg_2\frac{4}{4}\right) = 0.24$$

$$\text{Gain}(S, \overset{\text{student}}{\cancel{\text{income}}}) = 0.94 - \frac{7}{14}\left(\frac{6}{7}\lg_2\frac{6}{7} + \frac{1}{7}\lg_2\frac{1}{7}\right) - \frac{7}{14}\left(\frac{3}{7}\log_2\frac{3}{7} + \frac{4}{7}\log_2\frac{4}{7}\right) = 0.15$$

$$\text{Gain}(S, income) = 0.94 - \frac{4}{14}\left(\frac{2}{4}\lg_2\frac{2}{4} + \frac{2}{4}\lg_2\frac{2}{4}\right) - \frac{6}{14}\left(\frac{4}{6}\log_2\frac{4}{6} + \frac{2}{6}\lg_2\frac{2}{6}\right) - \frac{5}{14}\left(\frac{3}{5}\log_2\frac{3}{5} + \frac{2}{5}\lg_2\frac{2}{5}\right) = 0.03$$

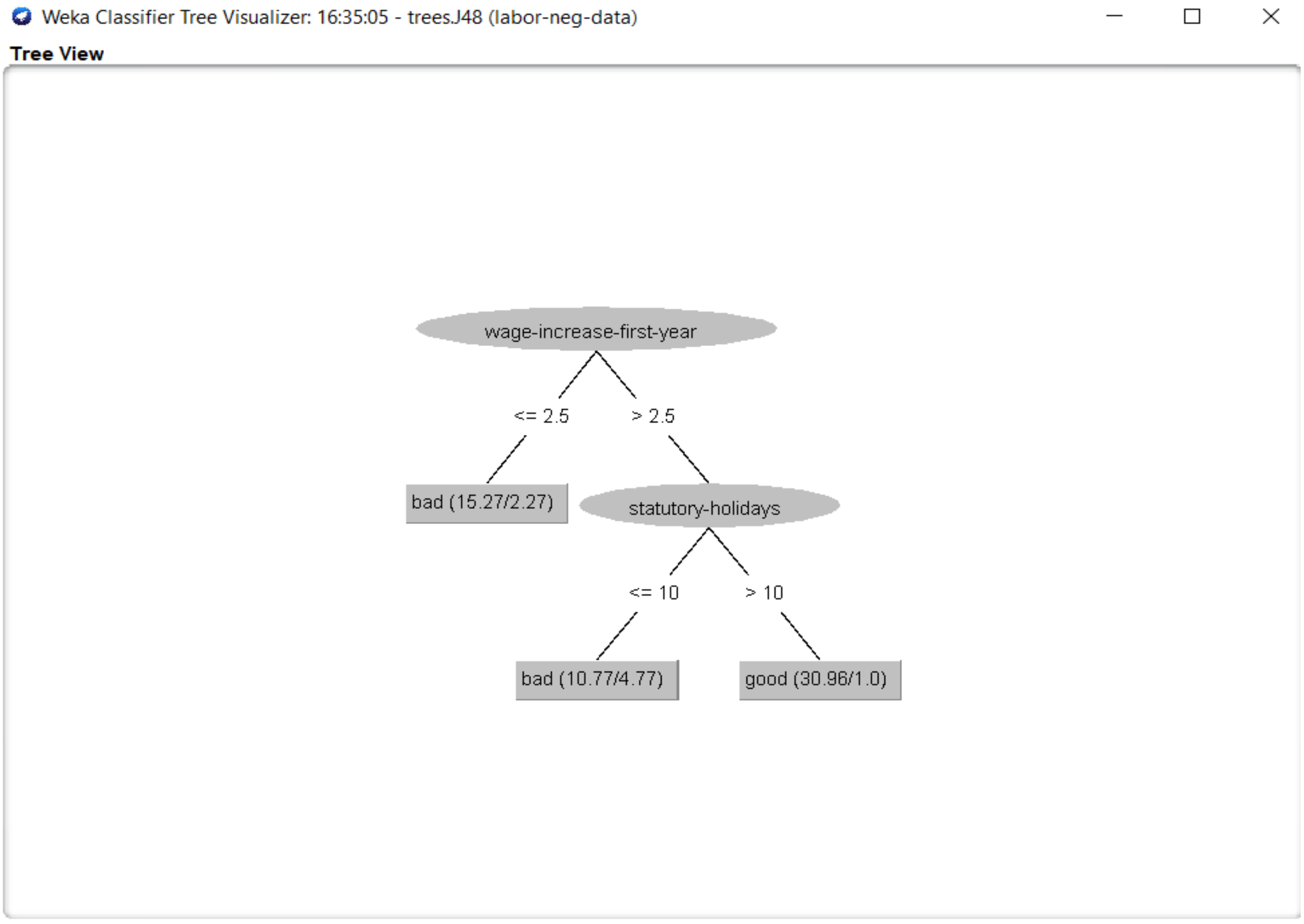
$$\text{Gain}(S, Credit) = 0.94 - \frac{8}{14}\left(\frac{6}{8}\lg_2\frac{6}{8} + \frac{2}{8}\lg_2\frac{2}{8}\right) - \frac{6}{14}\left(\frac{3}{6}\lg_2\frac{3}{6} + \frac{3}{6}\lg_2\frac{3}{6}\right) = 0.045$$

ریشه درخت Age است



* با دسته‌بندی Buy, Age
می‌توان لایه‌های بعدی درخت را
گسترش و یا می‌توانستیم Gain هر Feature مجدداً حساب کنیم

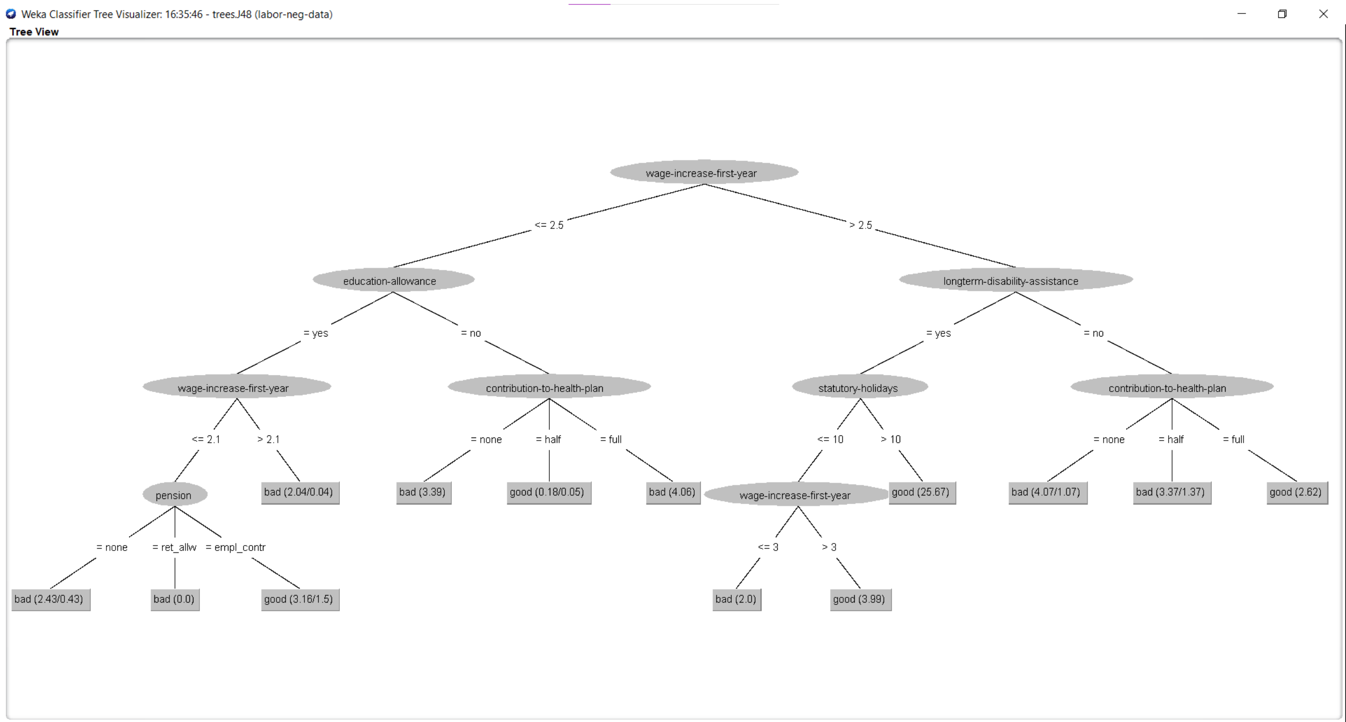
5-الف)



ب)

```
=== Summary ===

Correctly Classified Instances          50               87.7193 %
Incorrectly Classified Instances         7               12.2807 %
Kappa statistic                          0.745
Mean absolute error                      0.195
Root mean squared error                  0.304
Relative absolute error                 42.6664 %
Root relative squared error             63.6959 %
Total Number of Instances               57

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.950    0.162    0.760      0.950    0.844      0.758    0.918     0.809     bad
                 0.838    0.050    0.969      0.838    0.899      0.758    0.918     0.933     good
Weighted Avg.    0.877    0.089    0.896      0.877    0.880      0.758    0.918     0.890

=== Confusion Matrix ===

  a  b   <-- classified as
 19  1 |  a = bad
  6 31 |  b = good
```

ج)



د) هرس پیچیدگی طبقه بندی کننده نهایی را کاهش می دهد و در نتیجه دقت چیش بینی را با کاهش بیش برازش بهبود می بخشد. هرس درخت تصمیم به جلوگیری ازبرازش بیش از حد داده های اموزشی کمک می کند تا مدل ما به خوبی با داده های دیده نشده تعمیم یابد. هرس رخت تصمیم به معنای حذف درخت فرعی است و ان را با یک گره برگ جایگزین می کنیم