1. We have 14 samples in total with 9 Hires (H) and 5 Not Hired (NH). Information for the two classes H and NH is I (H,NH) = I (9,5) =$-\frac{9}{14}\log_2\frac{9}{14}-\frac{5}{14}\log_2\frac{5}{14}$= 0.94

   a. 130 has 5 samples, with 3H, 2 NH:

   $$I(3,2) = -\frac{3}{5}\log_2\frac{3}{5}-\frac{2}{5}\log_2\frac{2}{5} = 0.971$$

   160 has 5 result samples also with 2H, 3 NH:

   $$I\ (2,3) = -\frac{2}{5}\log_2\frac{2}{5}-\frac{3}{5}\log_2\frac{3}{5} = 0.971$$

   190 has 4 samples with 4H, 0NH:

   $$I(4,0) = -\frac{4}{4}\log_2\frac{4}{4} = 0$$

   Entropy for attribute Test Score (TS) is:

   $$E(TS) = \frac{5}{14}I(3,2) + \frac{5}{14}I(2,3) + \frac{4}{14}I(4,0) = 2 * 0.357 * 0.971 + 0.286 * 0 = 0.694$$

   Gain (TS) = I (9, 5) – E (TS) = 0.94 – 0.694 =0. 246

   Relevant Skills (RS) is a node of node 130 test score and has 2 classes yes (Y) and no (N) with 5 samples from 130 test scores.

   Yes has 3 samples with all 3 Hired while No has 2 with all not hired:

   $$I(3,0) = -\frac{3}{3}\log_2\frac{3}{3} = 0$$

   $$I(0,2) = -\frac{2}{2}\log_2\frac{2}{2} = 0$$

   Entropy for attribute Relevant skills is:

   $$E(RS) = \frac{3}{5}I(3,0) + \frac{2}{5}I(0,2) = 0$$

   Gain (RS) = I (3, 2) – E (RS) = 0.971

   b. 2 Hired and 2 Unidentified.

2. Probability of having cancer in the entire population is P (C) = 0.008

   Probability of not having cancer in the entire population is P (NC) = 1- P(C) = 0.992

   Probability of positive cancer test result and have cancer(True Positive) is P (P|C) = 0.98

   Probability having negative test result of having cancer but (False Negative) is P (N|C) = 0.02

   Probability of negative cancer test result and not have cancer (True Negative) is P (N|NC) = 0.97

Probability of having positive test result but not having cancer (False Positive) is P (P|NC) = 0.03

According to Bayes' Rules we have P(H|X) = $\frac{P(X|H)P(H)}{P(X)}$ with P(X)= P(x|H=1)P(H=1) + P(x|H=0)P(H=0)

a. The probability that the patient has cancer is:

$$P(C|P) = \frac{P(P|C)P(C)}{P(C)P(P|C) + P(NC)P(P|NC)} = \frac{0.98 * 0.008}{0.98 * 0.008 + 0.992 * 0.03} = 0.2$$

b. The probability that the patient doesn't have cancer is:

$$P(NC|P) = \frac{P(P|NC)P(NC)}{P(C)P(P|C) + P(NC)P(P|NC)} = \frac{0.03 * 0.992}{0.98 * 0.008 + 0.992 * 0.03} = 0.79$$

c. The diagnosis should be the patient most likely don't have cancer

3. P(Red|Y es) = 3 + 3 ∗ .5 5 + 3 = .56

   P(Red|No) = 2 + 3 ∗ .5 5 + 3 = .43

   P(SUV |Y es) = 1 + 3 ∗ .5 5 + 3 = .31

   P(SUV |No) = 3 + 3 ∗ .5 5 + 3 = .56

   P(Domestic|Y es) = 2 + 3 ∗ .5 5 + 3 = .43

   P(Domestic|No) = 3 + 3 ∗ .5 5 + 3 = .56

   P(Yes) * P(Red | Yes) * P(SUV | Yes) * P(Domestic|Yes) = .5 * .56 * .31 * .43 = .037

   P(No) * P(Red | No) * P(SUV | No) * P (Domestic | No) = .5 * .43 * .56 * .56 = .069

   Since 0.069 > 0.037, our example gets classified as 'NO'

4. We have sales as y and advertising dollars as x.

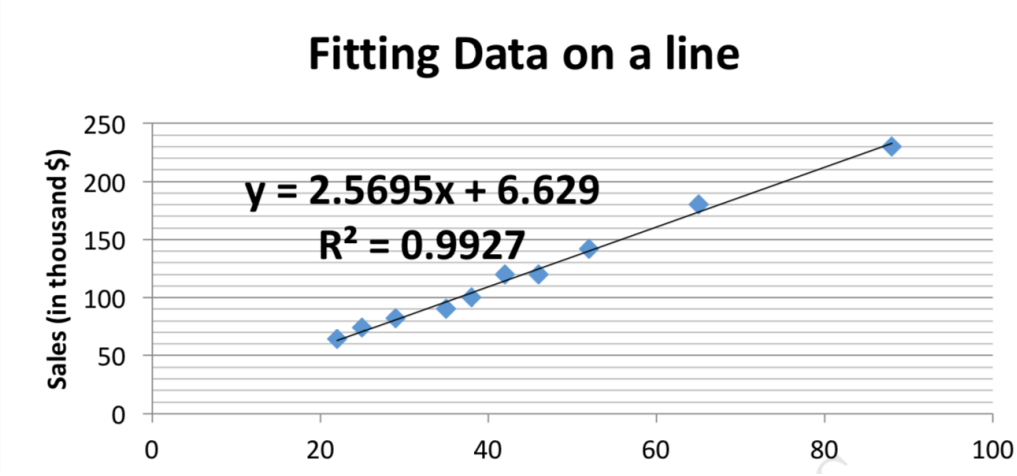
	AD(x)	S (y)	x-avg(x)	y-avg(y)	$(x\text{-avg}(x))^2$	$(y\text{-avg}(y))^2$	**Sum((x - avg(x))*(y-avg(y)))**
1	22	64	-22.2	-56.2	492.84		1247.64
2	25	74	-19.2	-46.2	368.64		887.05
3	29	82	-15.2	-38.2	231.04		580.64
4	35	90	-9.2	-30.2	84.64		277.84
5	38	100	-6.2	-20.2	38.44		125.24
6	42	120	-2.2	-0.2	4.84		0.44
7	46	120	1.8	-0.2	3.24		-0.36
8	52	142	7.8	21.8	60.84		170.04

9	65	180	20.8	59.8	432.64		1243.84
10	88	230	43.8	109.8	1918.44		4809.24
Sum	442	1202	0	20.2	3635.6		9341.6
Average	44.2	120.2					

a. The linear regression equation is written as: $y=b_0+b_1x$

$b_1 = \Sigma [(x_i - x)(y_i - y)] / \Sigma [(x_i - x)^2] = 9341.6/3635.6 = 2.5695$

$b_0 = y - b_1 * x = 120.2 - 2.5695(44.2) = 6.629$



b. The slope of the line is positive, which means that the sales increases with the increase of ad money.

c. $R^2 = 0.992$. Strong correlation and that the linear regression fits our data well. This means that we can use the model to predict the sales with ad money as a predictor.

d. Y=2.5695x+6.629 with x = 50K so y= 135.104K