1) a) $z = Xw + b$ $\quad X = [x_1, x_2]$

$= \beta_1 x_1 + \beta_2 x_2 + \beta_0$

$\sigma(z) = \frac{1}{1+e^{-z}}$ (sigmoid function)

$y\_pred = \sigma(\beta_0 + \beta_1 x_1 + \beta_2 x_2)$

optimisation using gradient descent

$\frac{\partial L}{\partial \beta_1} = \frac{\sum x_1 (y\_pred - y\_acc)}{n_1 + n_2}$

$\frac{\partial L}{\partial \beta_1} = \frac{\sum x_2 (y\_pred - y\_acc)}{n_1 + n_2}$

$\frac{\partial L}{\partial \beta_0} = \frac{\sum (y\_pred - y\_acc)}{n_1 + n_2}$

where L = loss fn
$n_1 + n_2$ = total no. of observations
y_acc = observed y value

$\beta_1 = \beta_1 - \alpha \frac{\partial L}{\partial \beta_1}$

$\beta_2 = \beta_2 - \alpha \frac{\partial L}{\partial \beta_2}$

$\beta_0 = \beta_0 - \alpha \frac{\partial L}{\partial \beta_0}$

initialize $\beta_0, \beta_1, \beta_2$ to 0 & apply this modification for a desired no. of iterations.
Here α = learning rate

b) likelihood function = $\prod_{i=1}^{n_1+n_2} (y_{pred})^{y_{acc}} (1-y_{pred})^{1-y_{acc}}$

log likelihood = $(n_1+n_2)[\, y_{acc} \log(y_{pred}) + (1-y_{acc}) \log(1-y_{pred})$

log loss = $-$(log likelihood)

2) a) $z = \beta_0 + \beta_1 x_1 + \beta_2 x_2$

$= -6 + (0.05)40 + 1(3.5)$

$= -6 + 2 + 3.5$

$= -0.5$

$p = \sigma(z) = \frac{1}{1+e^{-y_2}} = 0.622$

b) $p = \frac{1}{2} = \frac{1}{1+e^{-z}}$

$e^{-z} = 1$

$e^{z} = 1$

$z = 0$

$\beta_0 + \beta_1 x_1 + \beta_2 x_2 = 0$

$-6 + 0.05 x_1 + 1(3.5) = 0$

$0.05 x_1 = 2.5$

$x_1 = 50$ hrs

---

3) given: variance $= \sigma^2 = 36$

$\mu_1 = 10$ $\quad Y = 1$ = company issued dividend

$\mu_0 = 0$ $\quad Y = 0$ = company didnot issue

$P(y=1) = 0.8$

$P(y=0) = 0.2$

$$P(Y=1 \mid X=4) = \frac{f_1(4)\, P(Y=1)}{f_1(4)\, P(Y=1) + f_0(4)\, P(Y=0)}$$

where f_1 & f_0 are PDFs respectively

$$f_k(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu_k)^2}{2\sigma^2}}$$

$$f_1(4) = \frac{1}{\sqrt{2\pi 36}} e^{-\frac{(4-10)^2}{2(36)}} = \frac{1}{6\sqrt{2\pi}} e^{-1/2}$$

$$f_0(4) = \frac{1}{6\sqrt{2\pi}} e^{-\frac{(4-0)^2}{2(36)}} = \frac{1}{6\sqrt{2\pi}} e^{-2/9}$$

$$P(Y=1 \mid X=4) = \frac{0.8e^{-0.5}}{0.8e^{-0.5} + 0.2e^{-0.22}}$$

$$= 0.752$$

---

1) Decision Tree

For split 1, we can use feature X1 with threshold value of 'a'.

If $x_1 > a$ then class = +1

If $x_1 < a$ then check condition for split 2

↳ If $x_2 > b$ then class = +1

else class = -1

---

2) A random forest is a ML model which makes predictions by combining decisions of many different decision trees

- It is called forest bc its a combination of a large no. of different trees.

Each tree is trained independently and simultaneously on different data sets and features.

For classification, the tree vote and majority class wins

For regression, the trees' outputs are averaged.

- It is called random because
  - ↳ bootstrapping: the data set is randomly distributed (with replacement) among the trees.
  - ↳ Each tree use random set of features of the data point rather than choosing the best feature
- A single decision tree can overfit to data set, is sensitive to small changes and is unstable.
  A random forest is better because the mistakes of each tree balances out and generalisation is better. It is more stable & accurate.

---

3) • Ensemble method is where multiple models are combined to make a single prediction from predictions.

- It improves accuracy bc different models capture diff patterns and combining these makes prediction better.
  It reduces overfitting, bc many models are averaged.
  It increases stability bc errors made in1model can be evened out/corrected by the other models.
- Yes, bc it uses multiple decision trees, each tree makes its own prediction which is then combined by voting or averaging.

4) 1)
- TP = 180
- FP = 70
- TN = 730
- FN = 20

2) $\text{Accuracy} = \frac{TP+TN}{TP+FP+TN+FN} = \frac{180+730}{1000} = \frac{910}{1000} = 0.91$

↳ total

$\text{Precision} = \frac{TP}{TP+FP} = \frac{180}{250} = 0.72$

↳ total predicted +ve

$\text{Recall (Sensitivity)} = \frac{TP}{TP+FN} = \frac{180}{200} = 0.9$

↳ total real +ve

$\text{Specificity} = \frac{TN}{TN+FP} = \frac{730}{800} = \frac{3.65}{4} = 0.9125$

↳ total real -ve

$\text{F1 score} = \frac{2 \times \text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} = \frac{2(0.72)(0.9)}{1.62} = 0.9$

$= \frac{8}{10} = 0.8$

- I would prioritize Recall because it is the fraction of true +ves to total actual +ves which inlcudes False -ves. Thus Recall should be v high if FN is very costly bc our goal is to reduce FN
- If threshold is lowered, model is more willing to predict +ve. Thus TP & FP both increase. bc more acc +ve would be predicted +ve, TPT & more acc -ve would be wrongly predicted +ve, FPT

- Yes bc accuracy = $\frac{TP + TN}{total}$

one can predict acc +ve very well but not acc -ves while other can predict acc -ve correctly but not much acc+ve so TP+TN of both can be same but they are very different.