Marco Reina 7066486
Samuele Serri 7069839

a)

In a logistic regression model, the function:

$$Y = \frac{e^{f(x)}}{1+e^{f(x)}}$$

it is used for mapping the linear predictor $f(x)$ into $[0,1]$ so the output can be interpreted as a probability.

b)

$$p(y_1,...,y_n|\beta_0,\beta_1,...,\beta_n,x_1,...,x_n) = \prod_{i:y_i=1}^{n} \frac{e^{f(x_i)}}{1+e^{f(x_i)}} \prod_{i:y_i=0}^{n} 1 - \frac{e^{f(x_i)}}{1+e^{f(x_i)}}$$

If we take the logarithm:

$$\sum_{i:y_i=1}^{n} log\left(\frac{e^{f(x_i)}}{1+e^{f(x_i)}}\right) + \sum_{i:y_i=0}^{n} log\left(1 - \frac{e^{f(x_i)}}{1+e^{f(x_i)}}\right) =$$

$$= \sum_{i:y_i=1}^{n} f(x_i) - \sum_{i:y_i=1}^{n} log(1+e^{f(x_i)}) - \sum_{i:y_i=0}^{n} log(1+e^{f(x_i)})$$

In order to estimate the parameters of the model we need to find the maximum of the likelihood function. Finding the maximum of the log-likelihood function is equivalent because the logarithmic transformation is monotonic and preserves maximums.

c)

For classification problems discriminative models return a function that defines strict boundaries between the classes. Whereas generative models return the probability ($p_g(x)$) of x belonging in the class g for each class $g \in G$. Generative models estimate their parameters by maximizing a likelihood function while discriminative models minimize a loss function.