Solution 1:

Logistic regression is a classification model, that estimates posterior probabilities $\pi(x)$ by linear functions in x. For a binary classification problem the model can be written as:

$$\hat{y} = 1 \quad \Leftrightarrow \quad \pi(x) = \frac{1}{1 + \exp(-x^T\theta)} \geq a$$

For the decision boundary we have to set $\pi(x) = a$. Solving this equation for x yields:

$$\begin{aligned}
&\frac{1}{1 + \exp(-x^T\theta)} = a \\
\Leftrightarrow\ & 1 + \exp(-x^T\theta) = a^{-1} \\
\Leftrightarrow\ & \exp(-x^T\theta) = a^{-1} - 1 \\
\Leftrightarrow\ & \exp(-x^T\theta) = a^{-1} - 1 \\
\Leftrightarrow\ & -x^T\theta = \log(a^{-1} - 1) \\
\Rightarrow\ & x^T\theta = -\log(a^{-1} - 1)
\end{aligned}$$

For $a = 0.5$ we get:

$$x^T\theta = -\log(0.5^{-1} - 1) = -\log(2 - 1) = -\log(1) = 0$$

Solution 2:
See R code **mlr_l_4.R**

Solution 3:

a) $\pi_1(x) = \frac{e^{\theta_1^T x}}{e^{\theta_1^T x} + e^{\theta_2^T x}}$

   $\pi_2(x) = \frac{e^{\theta_2^T x}}{e^{\theta_1^T x} + e^{\theta_2^T x}}$

   $\pi_1(x) = \frac{1}{(e^{\theta_1^T x} + e^{\theta_2^T x})/e^{\theta_1^T x}} = \frac{1}{1 + e^{\theta^T x}}$ where $\theta = \theta_2 - \theta_1$ and $\pi_2(x) = 1 - \pi_1(x)$

b) Likelihood for one instance

$$L_i = P(y^{(i)} = k|\theta) = \prod_k \pi_k(x)^{I_k}$$

   where

$$\pi_k(x) = \frac{e^{\theta_k^T x}}{\sum_j e^{\theta_j^T x}}$$

$$\sum_{k=1}^{g} \pi_k = 1 \quad \text{and} \quad I_k = [y = k]$$

$$\sum_{k=1}^{g} I_k = 1$$

negative log likelihood for one instance:

$$-\log(L_i) = -\sum_{k=1}^{g} I_k \log(\pi_k(x))$$

negative log likelihood:

$$-\log(\mathcal{L}) = \sum_{i=1}^{n} -\log(L_i)$$

c) For g classes we get $g-1$ discriminant functions from the softmax $\pi_1(x), \ldots, \pi_{g-1}(x)$ which can be interpreted as probability. The probability for class g can be calculated by using $1 - \sum_{k=1}^{g-1} \pi_k(x)$. To estimate the class we are using majority vote:

$$\hat{y} = \operatorname*{arg\,max}_k \pi_k(x)$$

The parameter of the softmax regression is defined as parameter matrix where each class has its own parameter vector θ_k, $k \in \{1, \ldots, g-1\}$:

$$\theta = [\theta_1, \ldots, \theta_{g-1}]$$

Solution 4:

a) We need:

$$P(\text{Banana} = \text{yes}),\ P(\text{Color} = \text{yellow}|\text{Banana} = \text{¸yes}),\ P(\text{Form} = \text{round}|\text{Banana} = \text{¸yes}),$$
$$P(\text{Origin} = \text{imported}|\text{Banana} = \text{¸yes})$$

Since the features are just categorical we assume them to be Bernoulli distributed. Therefore, estimating the conditional probabilities is just calculating the relative frequencies conditioned on yes:

$$\begin{aligned}
&P(\text{Banana}) = 3/8 \\
&P(\text{Color} = \text{yellow}|\text{Banana} = \text{¸yes}) = 1/3 \\
&P(\text{Form} = \text{round}|\text{Banana} = \text{¸yes}) = 1/3 \\
&P(\text{Origin} = \text{imported}|\text{Banana} = \text{¸yes}) = 1
\end{aligned}$$

With the independence assumption of naive Bayes, the final step is to calculate the product of the probabilities above:

$$\begin{aligned}
&P(\text{Banana}|\text{Color} = \text{yellow},\ \text{Form} = \text{round},\ \text{Origin} = \text{imported}) \\
&\quad = P(\text{Banana}) \cdot P(\text{Color} = \text{yellow}|\text{Banana} = \text{¸yes}) \cdot P(\text{Form} = \text{round}|\text{Banana} = \text{¸yes}) \\
&\quad \cdot P(\text{Origin} = \text{imported}|\text{Banana} = \text{¸yes}) \\
&\quad = 3/8 \cdot 1/3 \cdot 1/3 \cdot 1 \\
&\quad = 1/24
\end{aligned}$$

b) For numerical features, we would use a different probability distribution, not the Bernoulli distribution. For example, for the information "Length" we could assume that $P(Length|yes)$ and $P(Length|no)$ both follow a normal distribution.