Exercise 1:

In logistic regression, we estimate the probability $\pi(\mathbf{x}) = \mathbb{P}(y = 1 \mid \mathbf{x})$. To decide if $\hat{y}$ is 0 or 1, we follow:

$$\hat{y} = 1 \quad \Leftrightarrow \quad \hat{\pi}(\mathbf{x}) \geq a$$

a) What happens if you are choosing $a = 0.5$? More precisely, from which value of $\theta^T x$ do you predict $\hat{y} = 1$ rather than $\hat{y} = 0$?

b) Explain (using words) why $a = 0.5$ is a sensible threshold.

Exercise 2:

Choose some of the classifiers already introduced in the lecture and visualize their decision boundaries for relevant hyperparameters. Use `mlbench::mlbench.spirals` to generate data and use `plot_learner_prediction` for visualization. To refresh your knowledge about `mlr3` you can take a look at `https://mlr3book.mlr-org.com/basics.html`.

Exercise 3:

a) What is the relationship between softmax $\pi_k(x) = \frac{\exp(\theta_k^T x)}{\sum_{j=1}^{g} \exp(\theta_j^T x)}$ and the logistic function $\pi(\mathbf{x}) = \frac{1}{1+\exp(\theta^T x)}$ for $g = 2$ (binary classification)?

b) The likelihood function of a multinomially distributed target variable with g target classes is given by

$$\mathcal{L}_i = \mathbb{P}(Y^{(i)} = y^{(i)} | x^{(i)}, \theta_1, \ldots, \theta_g) = \prod_{j=1}^{g} \pi_j(x^{(i)})^{\mathbb{1}_{\{y^{(i)}=j\}}}$$

where the posterior class probablities $\pi_1(x), \ldots, \pi_g(x)$ are modeled with softmax regression. Derive the likelihood function of n such independent target variables. How can you transform this likelihood function into an empirical risk function?
Hints:

- By following the maximum likelihood principle, we should look for parameters $\theta_1, \ldots, \theta_g$, which maximize the likelihood function.
- The expressions $\prod \mathcal{L}_i$ and $\log \prod \mathcal{L}_i$ (if this expression is defined) are maximized by the same parameters.
- The empirical risk is a *sum* of loss function values, not a *product*.
- Minimizing a scalar function multiplied with -1 is equivalent to maximizing the original function.

State the associated loss function.

c) Explain how the predictions of softmax regression (multiclass classification) looks like (probabilities and classes) and define the parameter space.

Exercise 4:

You are given the following table with the target variable `Banana`:

ID	Color	Form	Origin	Banana ?
1	yellow	oblong	imported	yes
2	yellow	round	domestic	no
3	yellow	oblong	imported	no
4	brown	oblong	imported	yes
5	brown	round	domestic	no
6	green	round	imported	yes
7	green	oblong	domestic	no
8	red	round	imported	no

a) We want to use a naive Bayes classifier to predict whether a new fruit is a Banana or not. Calculate the posterior probability $\pi(x)$ for a new observation (yellow, round, imported). How would you classify the object?

b) Assume you have an additional feature "Length", which measures the length in cm. Describe in 1-2 sentences how you would handle this numeric feature with Naive Bayes.