Exercise Sheet 1

Exercise

The equation

$$R(f_n) - R(f^*) = (R(f_n) - R(f_{\mathcal{H}})) + (R(f_{\mathcal{H}}) - R(f^*))$$

decomposes the term $R(f_n) - R(f^*)$ for Bayes-consistency as a sum of the estimation error $R(f_n) - R(f_{\mathcal{H}})$ and approximation error $R(f_{\mathcal{H}}) - R(f^*)$.

Task: Implement a classification problem using synthetic data points in $\mathbb{R}^2$ to illustrate the concepts of estimation and approximation error.

Details:

1. **Data Generation:** Generate synthetic data points in $\mathbb{R}^2$ from two classes.
2. **Hypothesis space:** Use the hypothesis class $\mathcal{H}$ of linear functions and apply logistic regression as learning algorithm.
3. **Risks:** Compute the empirical risks $R_n(f)$ and approximate the true risks $R(f)$ for $f\{f_n, f_{\mathcal{H}}, f = f^*\}$. Use a large sample size to approximate the true risks.
4. **Decomposition:** Demonstrate the estimation and approximation error for various sizes n of the training set. Repeat your experiment multiple times for each n.