

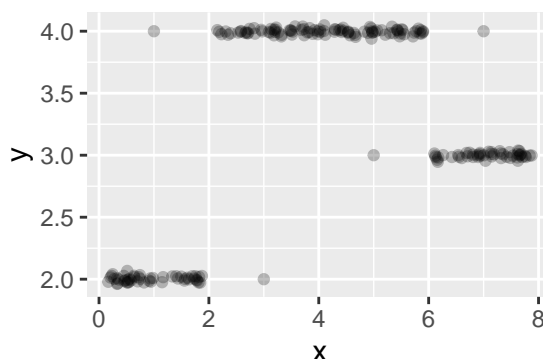
### Exercise 1: Naive Bayes

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. Estimate the posterior probability  $\hat{\pi}(\mathbf{x}_*)$  for a new observation  $\mathbf{x}_* = (\text{yellow}, \text{round}, \text{imported})$ . How would you classify the object?
- b) Assume you have an additional feature **Length** that measures the length in cm. Describe in 1-2 sentences how you would handle this numeric feature with Naive Bayes.

### Exercise 2: Discriminant Analysis



The above plot shows  $\mathcal{D} = ((\mathbf{x}^{(1)}, y^{(1)}), \dots, (\mathbf{x}^{(n)}, y^{(n)}))$ , a data set with  $n = 200$  observations of a continuous target variable  $y$  and a continuous, 1-dimensional feature variable  $\mathbf{x}$ . In the following, we aim at predicting  $y$  with a machine learning model that takes  $\mathbf{x}$  as input.

- a) To prepare the data for classification, we categorize the target variable  $y$  in 3 classes and call the transformed target variable  $z$ , as follows:

$$z^{(i)} = \begin{cases} 1, & y^{(i)} \in (-\infty, 2.5] \\ 2, & y^{(i)} \in (2.5, 3.5] \\ 3, & y^{(i)} \in (3.5, \infty) \end{cases}$$

Now we can apply quadratic discriminant analysis (QDA):

- i) Estimate the class means  $\mu_k = \mathbb{E}(\mathbf{x}|z = k)$  for each of the three classes  $k \in \{1, 2, 3\}$  visually from the plot. Do not overcomplicate this, a rough estimate is sufficient here.
- ii) Make a plot that visualizes the different estimated densities per class.

- iii) How would your plot from ii) change if we used linear discriminant analysis (LDA) instead of QDA? Explain your answer.
- iv) Why is QDA preferable over LDA for this data?
- b) Given are two new observations  $\mathbf{x}_{*1} = -10$  and  $\mathbf{x}_{*2} = 7$ . State the prediction for QDA and explain how you arrive there.

### Exercise 3: Decision Boundaries for Classification Learners

We will now visualize how well different learners classify the three-class `mlbench::mlbench.cassini` data set. Generate 1000 points from `cassini` using R or import `cassini_data.csv` in Python. Then, perturb the `x.2` dimension with Gaussian noise (mean 0, standard deviation 0.5), and consider the classifiers already introduced in the lecture:

- LDA,
- QDA, and
- Naive Bayes.

Plot the learners' decision boundaries. Can you spot differences in separation ability?

(Note that logistic regression cannot handle more than two classes and is therefore not listed here.)