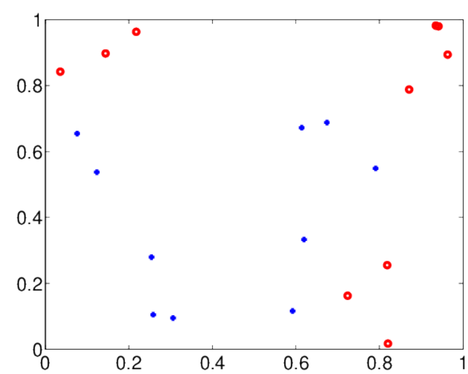
STAT 535 Homework 1
Out October 5, 2023
Due October 12, 2023, at noon

©Marina Meilă
mmp@stat.washington.edu

Problem 1 – Decision regions

a. Draw a reasonable decision tree for the following data (i.e. perform the algorithm "by hand").



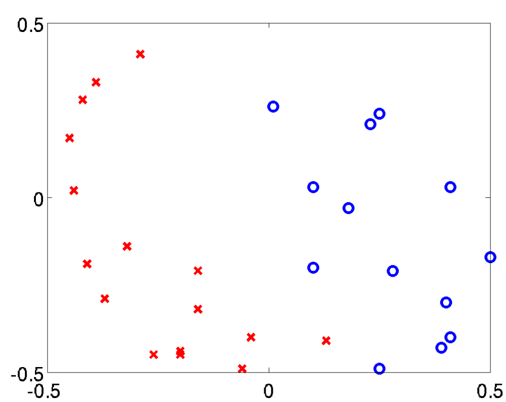
The data and file are provided on the Assignments web page.

b. Make a plot of the decision regions of the 1-NN classifiers for the same data.

c. Make a plot of the decision regions of the 3-NN classifiers for the same data. *(Make a plot as good as you can by hand)*

d. Draw an approximate plot of the decision region that LDA would obtain on the data below. Show the means of the two classes on the plot.

e. Draw an approximate plot of the decision region that QDA would obtain on the data in **e**.

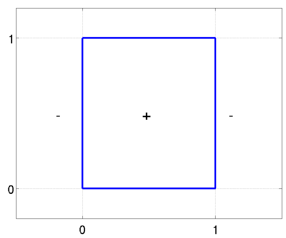


Note: it is OK to also run a program to do these exercises. But I recommend to first try doing what you can by hand. It would be a better practice for the quiz.

Problem 2 – More decision regions

Consider the decision regions below. You will try to obtain this decision boundary with a variety of classifier families, and see if it's possible exactly, approximately, or not at all. *The figure below is in* `fig_square_decregion.png`.

Decision regions



Classifier families

- ☐ decision tree (labeled drawing)
- ☐ decision tree with at most 2 levels (labeled drawing)
- ☐ 1-NN classifier (labeled drawing of a set of points and locations)
- ☐ 3-NN classifer (labeled drawing of a set of points and locations)
- ☐ Nadaraya-Watson with Gaussian kernel (labeled drawing of a set of points and locations)
- ☐ linear classifier (labeled drawing or formula)
- ☐ quadratic classifier (formula with numerical values or labeled drawing)

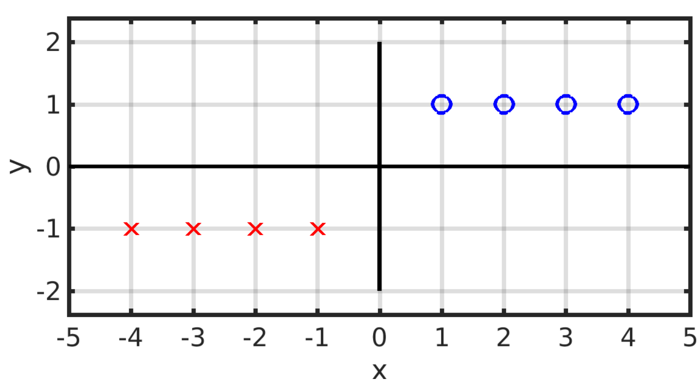
More precisely, for each of the predictor classes above

EITHER find a classifier in this class which realizes this decision region exactly
OR find a classifier in this class which approximates this decision region arbitrarily closely; be sure to say which parameter of the classifier controls how accurate is the approximation
OR give a brief argument why it is impossible to approximate this decision region

Problem 3 – Linear and logistic regression

Note: it is OK to also run a program to do these exercises. But I recommend to first try doing what you can by hand. It would be a better practice. We grade the answers on the assumptions that you did not have a computer available but that you understood the properties of the regressors.

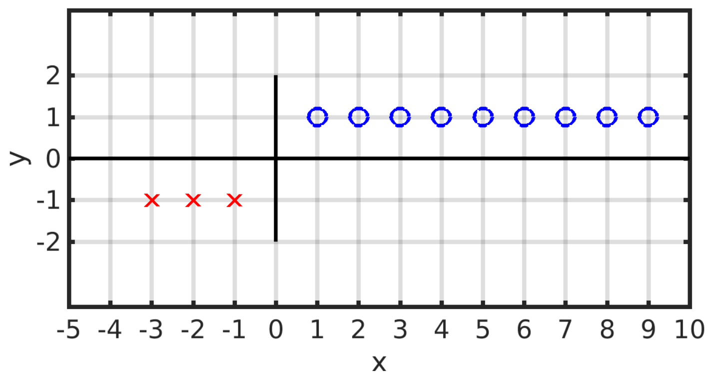
a. On the data below, draw approximately $p(x) = P_{Y=1|X}$ for a logistic regression classifier $f(x) = \beta x$ with $\beta = 1$ (see course notes for meaning of f.



b. On the same plot, draw approximately $p(x) = P_{Y=1|X}$ for a logistic regression classifier $f(x) = \beta x$ with $\beta = 0.1$.

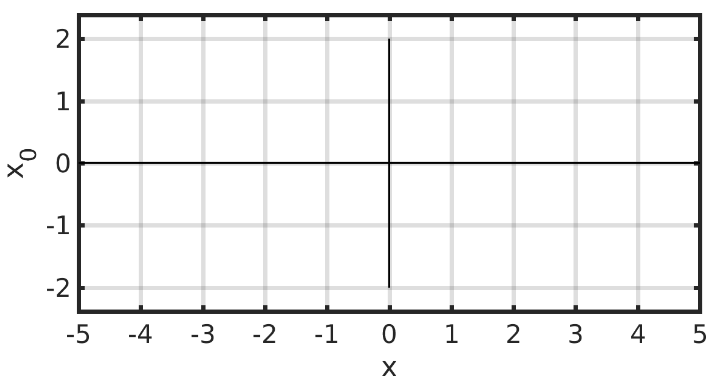
c. Suppose that you have the same data set, but you only have a Linear Regression program available. Draw what a linear regression function $f(x) = \beta x + \beta_0$ trained on these data will look like. Then mark the decision regions on the X axis.

d. Given the good experience with Linear Regression in **c.**, you decide to use Linear Regression on the new data set below. Draw what a linear regression function $f(x) = \beta x + \beta_0$ trained on these data will look like. Then mark the decision regions on the X axis.



e. Examine the classifier you obtained in **d**. Does it classify all the examples correctly? Explain in a sentence or two why or why not.

f. Let us return to the data from **a**. Represent these data points as points in $\mathbb{R}^2$ of the form $\tilde{x} = [x\, 1]^T$ on the graph below.



g. On the same graph above, draw the vector $\beta = [1\ -1]^T$ and the decision boundary of the linear classifier $\text{sgn}\beta^T\tilde{x}$.

h. For the data represented as in **f.** draw the centers $\mu_\pm$ and the decision region for the LDA classifier.

[Problem 4 – Basis function predictors – Optional, not graded]

Let $\mathcal{B} = \{b(\,,\xi),\ \xi \in \Xi\}$ be a finite or infinite *dictionary* and let the class of predictors $\mathcal{F}$ consist of finite linear combinations of dictionary functions, i.e. $\mathcal{F} \subseteq \{f(x) = \sum_{i=1}^{M} \beta_i b(x;\xi_i),\ \xi_{1:M} \in \Xi, \beta_{1:M} \in \mathbb{R}\}$. We say that $\mathcal{F}$ is a *basis function predictor*.

Show that each of the predictor classes below can be represented as a basis function classifier by finding a suitable dictionary and explaining what the coefficients β_i should be. For each classifier below, answer if $\mathcal{F}$ contains all the possible M-terms linear combinations over the dictionary $\mathcal{B}$ or a strict subset of them.

1. Regression trees with 1 level
2. Regression trees with any number of levels
3. 2 layer neural networks with linear outputs
4. Naive Bayes classifiers for binary classification, where $P_{X_j|y} = \mathcal{N}(\mu_{jy}, \sigma_j^2)$