Deadline: Friday November 20th , 10:00 Upload a `.pdf` file via LearnWeb. (e.g. exported Jupyter notebook)

1 Logistic Regression (12 points)

Given the toy dataset provided in Table ??,

A [3p] Perform 1 iteration gradient ascent.
(initial value $\beta^{(0)} = \big(-0.5,\ 1,\ 0\big)$ and learn-rate $\eta = 1$)

B [3p] Perform 1 iteration of Newtons method.
(initial value $\beta^{(0)} = \big(-0.5,\ 1,\ 0\big)$ and learn-rate $\eta = 1$)

x	y	class
-1	0	A
1	-1	A
0	1	B

Table 1

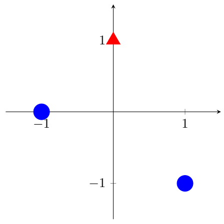

Figure 1

C [4p] In both cases, draw the decision boundaries and compute the loss terms before and after doing the optimization steps. What do you notice?

[2] What values would the parameters attain if we did an infinite number of steps?

2 Discriminant Analysis (10 points)

Given the toy dataset provided in Table ??,

A [4p] Compute the mean and covariance matrix for both the data points from class A and B.

B [3p] Predict whether the unlabeled datapoint belongs to class A and B using LDA

C [3p] Predict whether the unlabeled datapoint belongs to class A and B using QDA

x	y	class
1	2	A
2	1	A
2	2	A
1	-1	B
-1	-1	B
0	-2	B
-1	0	?

Table 2

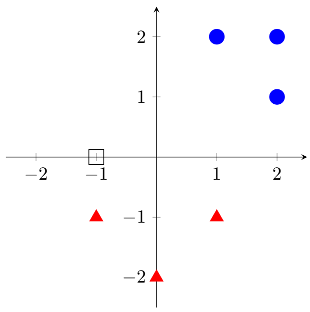

Figure 2

[8]Bonus Problem We consider the (in-)famous XOR dataset (Table ??), i.e. the truth table that describes the binary ”exclusive or” logic gate. Since the output is discrete we can consider it as a classification problem.

[2] Prove that the maximum accuracy a linear classifier can achieve on the XOR dataset is 75%. (a classifier is considered linear if its decision boundary is a hyperplane, i.e. of the form $H = \{x \mid \beta^{\mathsf{T}} x + \beta_0 = 0\}$)

[1] Prove that the loss function of logistic regression, $\log L^{\text{cond}}_{\mathcal{D}}$ is a concave function.

[4] Prove that for the XOR dataset, the loss function attains its global optimum at $\beta = 0$.

[1] Does this mean logistic regression is broken?

x_1	x_2	y
0	0	0
0	1	1
1	0	1
1	1	0

Table 3

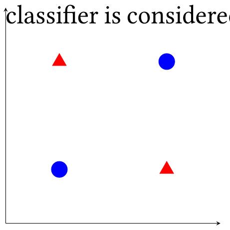

Figure 3