Exercise on binary classification

a. Generate a data set composed of 1000 samples according to the following model. The hypothesis $Y \in \{0,1\}$ is uniformly distributed, i.e., $\mathbb{P}[Y=0]=1/2$. The features X are scalar, with Gaussian distribution given the hypothesis, namely,

$$\ell(x|Y=0)=\frac{1}{2\pi}\exp\left\{-\frac{x^2}{2}\right\},\quad \ell(x|Y=1)=\frac{1}{2\pi}\exp\left\{-\frac{(x-1)^2}{2}\right\}.$$

(You can use the function "choice" in the "random" package of numpy to generate realizations of Y).

b. Implement the MAP classifier and compute its empirical error probability with a Monte Carlo simulation over the samples generated at point **a**. Compute the theoretical error probability of the MAP classifier and compare it against the empirical one.

c. Make a graphical representation of the posterior distribution under the two hypotheses, and highlight the decision threshold. Comment on what happens when the decision threshold changes by choosing different values of the problem parameters (try to change the means, standard deviation, prior probabilities).

d. Implement the classifier using the Neyman-Pearson criterion. Compute the decision threshold that corresponds to a false-positive rate $\alpha=0.1$. Using the chosen threshold, compute the error probabilities with a Monte Carlo simulation over the samples generated at point **a**. (To evaluate the Q function and its inverse, you can use "scipy.stats.norm.cdf", "scipy.stats.norm.ppf", "scipy.stats.norm.sf", and "scipy.stats.norm.isf").

e. Represent graphically the ROC for the Neyman-Pearson classifier. Compare the ROC obtained by evaluating empirically the true-positive and false-positive rates and the curve obtained with the theoretical formulas illustrated during the course. Comment on what happens when the parameters of the problem change (means and standard deviation).

Exercise on the stochastic gradient descent algorithm for supervised classification

a. Generate a training set with $N_{\text{train}}=5000$ samples corresponding to the following binary classification problem. The hypothesis $Y\in\{-1,+1\}$ is uniformly distributed. The features $X\in\mathbb{R}^2$ have independent components X_1 and X_2. The distribution of the features given the hypothesis Y is:

$$\ell(x_1|Y=-1)=\frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{(x_1-0.9)^2}{2}\right\},\qquad \ell(x_2|Y=+1)=\frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{(x_2-0.9)^2}{2}\right\},$$

$$\ell(x_1|Y=-1)=\frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{(x_1+0.85)^2}{2}\right\},\qquad \ell(x_2|Y=+1)=\frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{(x_2+0.85)^2}{2}\right\}.$$

Represent with a scatter plot the training data and comment on whether a logistic classifier is useful for this classification problem.

b. Write a function that computes the stochastic gradient of the cost function used to estimate the parameters of a logistic classifier.

c. Implement the stochastic gradient descent algorithm. The implementation should store the parameters estimated across the iterations of the algorithm, and should work both with constant and diminishing step-size.

d. Apply your implementation of the SGD algorithm to the training set generated at point **a** to estimate the parameters of a logistic classifier. Comment on the convergence properties of the SGD algorithm by evaluating the evolution of the parameters over time. Try different choices for the step-size, under both settings with constant or diminishing step-size.

e. Derive the expression of the MAP classifier for the assigned problem. Does the MAP classifier belong to the family of classifiers described by the logistic model? Which are the optimal parameters the SGD algorithm should converge to?

f. Compute the empirical error probability and the empirical accuracy of the logistic classifier on a test set with $N_{\text{test}}=1000$ samples.

g. Represent graphically the test data and depict the decision regions used by the logistic classifier.

Exercise on the stochastic gradient descent algorithm for supervised linear regression.

a. Generate a training set with $N_{\text{train}}=5000$ samples corresponding to a linear regression problem

$$Y=\beta_0+\beta_1X+\mathcal{E},$$

where X is a standard Gaussian, whereas $\mathcal{E}$ is a zero-mean Gaussian (independent of X) with variance 0.1. Choose freely the values of β_0 and β_1.

b. Write a function that computes the stochastic gradient of the cost function used to estimate β_0 and β_1.

c. Apply your implementation of the SGD algorithm to the training set generated at point **a** to estimate β_0 and β_1. Comment on the convergence properties of the SGD algorithm by evaluating the evolution of the parameters over time. Try different choices for the step-size, under both settings with constant or diminishing step-size. Is the algorithm able to estimate correctly the parameters?