Problem 1

Create a gradescope account on https://gradescope.com. Use course entry code 97XV43 and use your PDX email and PDX Student ID and your given name and family name. Turning this assignment in via gradescope will be confirmation that you have completed this problem.

Problem 2

Set up your Python environment. To complete this week’s assignment, you will need to install the following:


  conda create --name your_env_name python=3.7
  conda activate your_env_name

  to establish and activate your environment. Within it, you can install the packages needed for this homework.

- *PyTorch* from https://pytorch.org/. If you are installing on a local machine, you can use the quick start widget, which will generate the conda command for your install, something like:

  conda install pytorch torchvision --c pytorch

- *Jupyter Notebook* from https://jupyter.org/. This is a browser based programming interface that is very helpful for iterative code development. Data scientists commonly use it. Alternatives are Google Colab, an IDE like PyCharm, or simply using a text editor and the command line. You can install it with:

  conda install jupyter

- matplotlib with detailed instructions here: https://matplotlib.org/users/installing.html. This should work:

  conda install matplotlib

That’s it! You should be ready to develop your first neural network ;-)
Problem 3 (Theory)

Prove that the PLA eventually converges to a linear separator for separable data. The following steps will guide you through the proof. Let \( w^* \) be an optimal set of weights. These weights separate the data. This proof will show that the PLA weights \( w(t) \) get ”more aligned” with \( w^* \) every iteration. For simplicity, assume \( w(0) = 0 \).

(a) Let \( \rho = \min_{1 \leq n \leq N} y_n(w^*^T x_n) \). Show that \( \rho > 0 \).

(b) Show that \( w^T(t)w^* \geq w^T(t-1)w^* + \rho \), and conclude that \( w^T(t)w^* \geq t\rho \). [Hint: Use induction.]

(c) Show that \( ||w(t)||^2 \leq ||w(t-1)||^2 + ||x(t-1)||^2 \).

[Hint: \( y(t-1)(w^T(t-1)x(t-1)) \leq 0 \) because \( x(t-1) \) was misclassified by \( w(t-1) \).]

(d) Show by induction that \( ||w(t)||^2 \leq tR^2 \), where \( R = \max_{1 \leq n \leq N} ||x_n|| \).

(e) Using (b) and (d), show that

\[
\frac{w^T(t)}{||w(t)||} w^* \geq \sqrt{t\frac{\rho}{R}},
\]

and hence prove that

\[
t \leq \frac{R^2 ||w^*||^2}{\rho^2}.
\]

[Hint: \( \frac{w^T(t)w^*}{||w(t)|| ||w^*||} \leq 1 \). Why?]

Problem 4 (Practice)

For this problem, you will extend my in-class tutorial to further experiment with perceptrons and the perceptron learning algorithm. You will analyze the algorithm with different datasets and dimensionality.

(a) Add the final hypothesis \( g \) to the existing data plot in the tutorial. Comment on whether \( f \) is close to \( g \).

(b) Repeat the original experiment with another randomly generated dataset of size 100. Report the number of updates the algorithm takes before converging.

(c) Repeat (b) with a dataset of size 1,000. Compare your results.

(d) Modify the algorithm so that it takes \( x_n \in \mathbb{R}^10 \) instead of \( \mathbb{R}^2 \). Randomly generate a linearly separable dataset of size 1,000 with \( x_n \in \mathbb{R}^10 \) and feed the dataset to the algorithm. How many updates does it take to converge?

(e) Repeat the algorithm on the same dataset as (d) for 100 experiments. In the iterations of each experiment, pick \( x(t) \) randomly (i.e., pick the next one used for the weight vector update randomly) instead of deterministically. Plot a histogram for the number of updates that the algorithm takes to converge.

(f) Summarize you conclusions with respect to accuracy and running time as a function of \( N \) and \( d \).

*** IMPORTANT: No hard copies accepted. Submit typed or written answers in PDF format for the theoretical portion. Submit answers to the programming questions in PDF format, along with plots. You may choose to markdown your Jupyter Notebook output. In addition, you must upload your source code files. ***