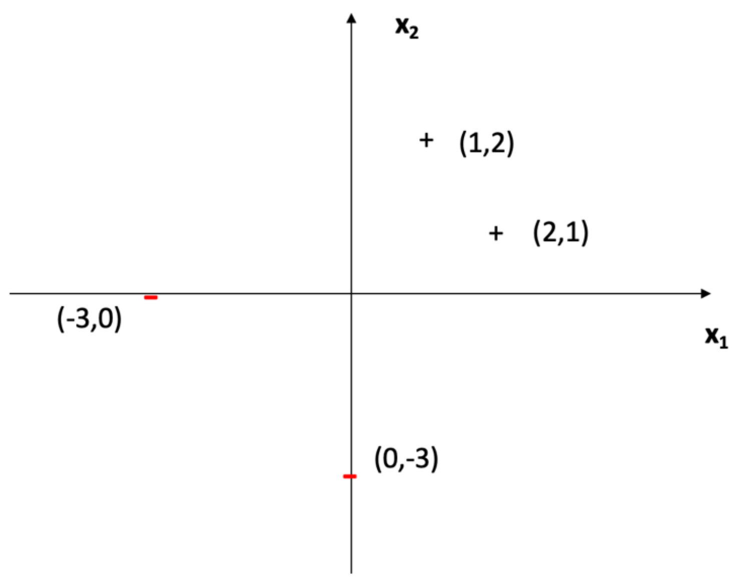
Justin Lam
77333383

CIVL 498A HW-4: SVM, Kernels, MLP.

Note: For hand calculation problems, you are supposed to write down your answers on a piece of paper and then submit the scanned version of the paper(s). Show your work by writing out derivations if there is any. For coding problems, you can either use Google Colab or Jupyter Notebook (on your own machine) to write and run the code. When submitting, submit only the ipynb file (you should use relative file directory for input files e.g., "./data.csv" in your code, so when I run the file on my local computer, files would load normally), and make sure when "run all" is clicked, all required results from the questions are shown (this is how I will grade your answer. If when I "run all" and your code crashes, it will be deemed as wrong).

1. **(Hand calculation)** You observe four data points on the x_1, x_2 plane. These four points belong to two classes, i.e., two positive examples (represented by "+"), and two negative examples (represented by "-").



Find the decision boundary that separates the two classes with the largest geometric margin.

1.1. Write the problem in the form of OP1, and derive OP2 from OP1.

1.2. Solve OP2 using Lagrange multiplier and find the decision boundary $w^T x+b = 0$

1.1

Points : $S = \{(1,2)\ (2,1)\ (-3,0)\ (0,-3)\}$

Decision Boundary $= w^T x^{(i)} + b$

$$\gamma^{(i)} = \text{distance} = \frac{w^T x^{(i)} + b}{||w||}$$

$$\gamma^{(i)} = \frac{y^{(i)} \cdot [w^T x^{(i)} + b]}{||w||} = \frac{\hat{\gamma}^{(i)}}{||w||}$$

For OP1: $\max \gamma$

$$\frac{y^{(i)} \cdot [w^T x^{(i)} + b]}{||w||} \geq \gamma$$

For OP 2: use invariant nature of geo. margin

Let $||w|| = \frac{1}{\gamma}$

$$\max_{w,b} = \frac{1}{||w||} = \min ||w|| = \min \frac{1}{2} ||w||^2$$

$$y^{(i)} [w^T x^{(i)} + b] \geq 1$$

1.2

Points: $S = \left\{ \overset{+}{(1,2)}\ \overset{+}{(2,1)}\ \overset{-}{(-3,0)}\ \overset{-}{(0,-3)} \right\}$

$$\min \frac{1}{2}\|w\|^2 \rightarrow 1 - y^{(i)}\left[w^T x^{(i)} + b\right] \leq 0$$

Lagrangian:

$$\max L(w, \alpha) = f(w) + \sum_{i=1}^{m} \alpha_i \cdot g_i(w)$$

$$= \frac{1}{2}\|w\|^2 + \sum_{i=1}^{n} \left[1 - y^{(i)}\left(w^T x^{(i)} + b\right)\right] \cdot \alpha_i$$

$$\max L(w, \alpha, b) = \frac{1}{2}\left(\sqrt{w_1^2 + w_2^2}\right)^2 + \alpha_1\left[1 - 1 \times (w_1 + 2w_2 + b)\right]$$
$$+ \alpha_2\left[1 - 1 \times (2w_1 + w_2 + b)\right]$$
$$+ \alpha_3\left[1 + 1 \times (-3w_1 + b)\right]$$
$$+ \alpha_4\left[1 + 1 \times (-3w_2 + b)\right]$$

$$\max L(w, \alpha, b) = \frac{1}{2}\left(w_1^2 + w_2^2\right) + \alpha_1 - \alpha_1(w_1 + 2w_2 + b)$$
$$+ \alpha_2 - \alpha_2(2w_1 + w_2 + b) + \alpha_3 + \alpha_3(-3w_1 + b)$$
$$+ \alpha_4 + \alpha_4(-3w_2 + b)$$

Find max $dL(w, \alpha, b)$:

$$\frac{dL}{dw_1} = \underset{1}{w_1} - \underset{4}{\alpha_1} - \underset{5}{2\alpha_2} - \underset{6}{3\alpha_3}$$

$$
\left.
\begin{aligned}
\frac{dL}{dw_2} &= \underset{2}{w_2} - 2\underset{4}{\alpha_1} - \underset{5}{\alpha_2} - 3\underset{7}{\alpha_4} \\
\frac{dL}{db} &= -\underset{4}{\alpha_1} - \underset{5}{\alpha_2} + \underset{6}{\alpha_3} + \underset{7}{\alpha_4} \\
\frac{dL}{d\alpha_1} &= 1 - (\underset{1}{w_1} + 2\underset{2}{w_2} + \underset{3}{b}) \\
\frac{dL}{d\alpha_2} &= 1 - (2\underset{1}{w_1} + \underset{2}{w_2} + \underset{3}{b}) \\
\frac{dL}{d\alpha_3} &= 1 + (-3\underset{1}{w_1} + \underset{3}{b}) \\
\frac{dL}{d\alpha_4} &= 1 + (-3\underset{2}{w_2} + \underset{3}{b})
\end{aligned}
\right\} = 0
$$

Plug in solver: Let $\underset{1}{w_1} - \underset{2}{w_2} - \underset{3}{b} - \underset{4}{\alpha_1} - \underset{5}{\alpha_2} - \underset{6}{\alpha_3} - \underset{7}{\alpha_4}$

$$
\begin{bmatrix} w_1 \\ w_2 \\ b \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \alpha_4 \end{bmatrix}
\rightarrow
\begin{bmatrix} 1/3 \\ 1/3 \\ 0 \\ 2/9 \\ -1/9 \\ 1/9 \\ 0 \end{bmatrix}
\rightarrow
$$

$$w^T x + b = 0$$

$$\tfrac{1}{3}x_1 + \tfrac{1}{3}x_2 + 0 = 0$$

$$\boxed{\tfrac{1}{3}x_1 + \tfrac{1}{3}x_2 = 0}$$

2. **(Hand calculation)** Find the mapping functions $\phi(x)$ and $\phi(z)$ for the kernel function K(x, z), such that $K(x, z) = (x^T z + c)^2$ where c is a constant. Assume $x = [x_1, x_2]^T$. Note, $\phi(x)$ and $\phi(z)$ have the same form, just with x substituted for z. Hint, $\phi(x)$ should be a 7 by 1 vector.

$$x = [x_1, x_2]^T = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \qquad \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \times [z_1, z_2]$$

$$K(x,z) = \left((x_1, x_2)^T \cdot (z_1, z_2) + c \right)^2$$

$$= (x_1 z_1 + x_2 z_2 + c)^2 = (x_1 z_1 + x_2 z_2 + c)(x_1 z_1 + x_2 z_2 + c)$$

$$= x_1^2 z_1^2 + x_1 x_2 z_1 z_2 + x_1 z_1 c + x_2 x_1 z_2 z_1 + x_2^2 z_2^2 + x_2 z_2 c + x_1 z_1 c + x_2 z_2 c + c^2$$

$$K(x,z) = \phi^T(x) \cdot \phi(z)$$

$$\phi(x) = \begin{bmatrix} x_1^2 \\ x_1 x_2 \\ \sqrt{2c}\, x_1 \\ x_2 x_1 \\ x_2^2 \\ \sqrt{2c}\, x_2 \\ c \end{bmatrix} \qquad \phi(z) = \begin{bmatrix} z_1^2 \\ z_1 z_2 \\ \sqrt{2c}\, z_1 \\ z_2 z_1 \\ z_2^2 \\ \sqrt{2c}\, z_2 \\ c \end{bmatrix}$$

3. **(Coding Problem)** Use the datasets named "train.csv" and "test.csv" for the following questions. This data is a partial dataset from the CIFAR 100 data, which contains images of 100 classes, categorized under 20 super classes. Refer to the following website for details of the full dataset (https://www.cs.toronto.edu/~kriz/cifar.html).

The provided partial datasets contains 2500 images for training (in the train.csv file) and 500 images for testing (in the test.csv file). The combined 3000 images are from the superclass "large man-made outdoor things", including classes of "bridge", "castle", "house", "road", "skyscraper". Each image is 32-pixel by 32-pixel, with RGB channels.

3.1. Read in the training data, use the first 3072 columns as the input X, and the final column as the output y. Note the y column uses numerical values instead of strings as the class name.

3.2. Use the SVC classifier from sklearn to classify the images into 1 of the 5 classes. Use 5-fold cross validation to evaluate and find the best model, from all combinations of the following parameters ("C" in {0.5, 1, 2}, "kernel" in {"linear", "poly", "rbf"}, when using the poly kernel, select "degree" from {2, 3, 5}). In total, there are 15 models to select from, each using one specific combination of the parameters. Use the average cross validation accuracy (sklearn.metrics.accuracy_score) to select the best performing model.

3.3. Use the selected model to test your performance on the test set and report the accuracy_score.

3.4. Do you think the accuracy_score is a good choice as the performance metric? If not, which one of the classification metrics from 3.3.1.1 classification section of this page (https://scikit-learn.org/stable/modules/model_evaluation.html) you think is a better metric and why? You can write out your answer as a comment in one of the cells.