ECE 09495/09595

Assignment 1_Jacob Matteo

Instructions

1. Max Credit: 100 Points
2. All questions are from the Textbook – Dive into Deep Learning (https://d2l.ai/).
3. Submit a single PDF.

Questions

1. **Part – 1 (10 Problems)**   **50 points**
   a. **Section 3.1.6 – Q 1, 2 and 3.**

   1)

   1) Assume we have data $x_1, \ldots, x_n \in \mathbb{R}$, Find b so $\sum_i (x_i - b)^2$ is minimized
   a) Find an analytic solution for b's optimal value
   b is like c in $y = mx + b$, $b \in \mathbb{R}^{2 \times 1}$, $b = $ bias $0 = (x_i - b)^2 \Rightarrow 0 = x_i - b \Rightarrow b = x_i$
   $w = (x^T x)^{-1} x^T y = (x^T y)^{-1} x^T \sum_i (x_i - b)^2$

   a.
   b. This function relates to the normal distribution as the closer b is to 0, the larger the value outputted by the function. This creates a normal distribution around b=0, shrinking as |b| tends towards |x|. However, b increases again as |b| increases beyond |x|.

   2)

   a.

   b.

   c.

   d.

   3)

   3) a) $p(\epsilon) = \frac{1}{2} e^{-|\epsilon|} \rightarrow -\log$ likelihood under $-\log P(y|x)$
   $J(y, \hat{y}) = -\log p(y, \hat{y}) = -y \log \hat{y} - (1-y)\log(1-\hat{y})$
   $P(y|x) = \frac{1}{2} \exp(-|y - w^T x - b|)$
   $-\log P(y|x) = +|y - w^T x - b| \log(\frac{1}{2}e) = (|y - w^T x - b| \log \frac{1}{2})$

   a.

   b.

c. In a stochastic gradient, this function iterates to a point in which the algorithm can no longer minimize loss and will bounce between a minimal-loss setting and a much higher-loss setting.

b. Section 3.2.9 – Q 5, and 7.

5) The reason the reshape function is needed in the square_loss function outputs a 1D array rather than an n by d matrix, so the reshape function is there to put the outputted values back in the correct locations

7) When data_iter's batch size does not evenly divide the number of examples, this causes unpredictable behavior which results in incorrect values as the function will land between entries (the location it ends up on isn't a whole number) so the function breaks down.

c. Section 3.4.9 – Q 1, 2, and 3.

1)

    a.

    b.

2)

    a.

    b.

3)

    a.
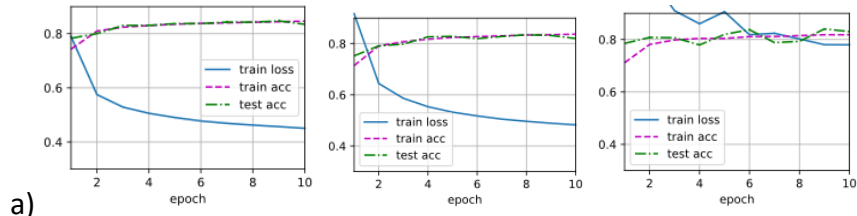
    b.

    c.

    d.

    e.

d. Section 3.6.8 – Q 4.

4) While usually you would want your algorithm to return the most-likely label, this is not always the case. For example, if you had an algorithm which separated types of mushrooms for you to eat, you would want to be absolutely sure the mushroom you pick isn't poisonous. Even 1% likely can be too high in circumstances such as this.

e. Section 3.7.6 – Q 1.

a)

    i.   What is the accuracy of your best model after hyperparameter tuning?
        Roughly 8 ¼ with stock settings except epochs # = 100.
   ii.   Provide a plot of epochs vs. train and test accuracy.
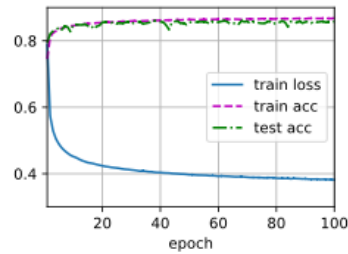
2. Part – II   50 points

   a. Complete your submission of Predicting House Prices on Kaggle – Section 4.10
      https://d2l.ai/chapter_multilayer-perceptrons/kaggle-house-
      price.html#predicting-house-prices-on-kaggle
   b. Attach the screenshot of your submission in the PDF file.

Your most recent submission

Name	Submitted	Wait time	Execution time	Score
submission.csv	just now	0 seconds	0 seconds	0.16696

Complete

Jump to your position on the leaderboard ▾