

CS229 Problem Set #1

[25 points] Locally Weighted Linear Regression

(a) [10 points]

Consider a linear regression problem in which we want to “weight” different training examples differently. Specifically, suppose we want to minimize:

$$J(\theta) = \frac{1}{2} \sum_{i=1}^m w^{(i)} \left(\theta^\top x^{(i)} - y^{(i)} \right)^2.$$

In class, we worked out what happens for the case where all the weights (the $w^{(i)}$'s) are the same. In this problem, we will generalize some of those ideas to the weighted setting.

i. [2 points] Show that $J(\theta)$ can also be written as:

$$J(\theta) = (X\theta - y)^\top W(X\theta - y),$$

for an appropriate matrix W , where X and y are as defined in class. Clearly specify each of the matrix W , X and vector y .

ii. [4 points] If all the $w^{(i)}$'s equal 1, then we saw in class that the normal equation is:

$$X^\top X \theta = X^\top y,$$

and that the value of θ that minimizes $J(\theta)$ is given by $(X^\top X)^{-1} X^\top y$. By finding the derivative $\nabla_\theta J(\theta)$ and setting it to zero, generalize the normal equation to this weighted setting, and give the new value of θ that minimizes $J(\theta)$ in closed form as a function of X , W , and y .

iii. [4 points] Suppose we have a dataset $\{(x^{(i)}, y^{(i)}); i = 1, \dots, m\}$ of m independent examples, but we model the $y^{(i)}$'s as drawn from conditional distributions with different levels of variance $\sigma^{(i)2}$. Specifically, assume the model:

$$p(y^{(i)} | x^{(i)}; \theta) = \frac{1}{\sqrt{2\pi\sigma^{(i)}}} \exp \left(-\frac{(y^{(i)} - \theta^\top x^{(i)})^2}{2\sigma^{(i)2}} \right).$$

Show that finding the maximum likelihood estimate of θ reduces to solving a weighted linear regression problem. State clearly what the $w^{(i)}$'s are in terms of the $\sigma^{(i)}$'s.

(b) [10 points]

Coding problem. We will now consider the following dataset (the formatting matches that of Datasets 1-4, except $x^{(i)}$ is 1-dimensional): `data/ds5{train, valid, test}.csv`.

In `src/p05b_lwr.py`, implement locally weighted linear regression using the normal equations you derived in Part (a) and using:

$$w^{(i)} = \exp\left(-\frac{\|x^{(i)} - x\|^2}{2\tau^2}\right).$$

Train your model on the `train` split using $\tau = 0.5$, then run your model on the `valid` split and report the mean squared error (MSE). Finally, plot your model's predictions on the validation set (plot the training set with blue 'x' markers and the validation set with red 'o' markers). Does the model seem to be under- or overfitting?

(c) [5 points]

Coding problem. We will now tune the hyperparameter τ . In `src/p05c_tau.py`, find the MSE value of your model on the validation set for each of the values of τ specified in the code. For each τ , plot your model's predictions on the validation set in the format described in part (b). Report the value of τ which achieves the lowest MSE on the `valid` split, and finally report the MSE on the `test` split using this τ -value.