

BLM462E Homework#2

Learning Curve & Parameter Selection

Deadline: April 16, 2020 Thursday 23:59

1) Learning Curve: In this part, you are required to plot the learning curve for regression problem:

- i. Visit https://github.com/ageron/handson-ml2/blob/master/04_training_linear_models.ipynb
- ii. You should invoke the `plot_learning_curve` with different models (Linear Regression, Linear Regression with polynomial features)
- iii. You can use any dataset with one feature.
- iv. Provide your discussion (Does adding more training data improve an overfitting model? Why?)

2) Parameter Selection: In this part, you are required to apply Ridge to boston dataset. You should use the polynomial features on the boston dataset. You must search over the degree parameters together with the parameter alpha of Ridge. You should

- i. Load boston dataset
- ii. Scale the data
- iii. Compute the polynomial features
- iv. Try the values 1, 2, 5, 10 for degree and 0.001, 0.01, 0.1, 1, 10, 100 for alpha.
- v. Use the `GridSearchCV` class and select the best parameter settings for ridge regression.
- vi. Apply Ridge regression with best setting and report the train and test results.

You are required to write a detailed report that includes your results and discussions. You should cite your references. Please zip and submit all your files (report and jupyter notebook files).

You may discuss an assignment and the general approach to a problem with your instructor, TA, or your classmates. However, the final submitted work should be totally yours.

Berna Kiraz.

BLM535 Ödev#2

Öğrenme Eğrisi & Parametre Seçme

Son tarih: 16 Nisan 2020 Perşembe 23:59

1) Öğrenme Eğrisi: Bu kısımda, regresyon problem için öğrenme eğrisi çizdirmeniz beklenmektedir:

- i. https://github.com/ageron/handson-ml2/blob/master/04_training_linear_models.ipynb sayfasını ziyaret edin.
- ii. `plot_learning_curve` metodunu farklı modellerle çağırın. (Linear Regression, Linear Regression with polynomial features)
- iii. Tek bir özneliği olan herhangi bir data kullanabilirsiniz.
- iv. Yorumlarınızı belirtiniz. (Daha fazla eğitim verisi eklemek aşırı öğrenmiş model için bir çözüm olmuş mudur? Neden?)

2) Parametre Seçimi: Bu kısımda boston veri kümesine Ridge yöntemini uygulayacaksınız. Boston veri kümesi üzerinde polinom öznelikleri kullanmanız gerekmektedir. Derece ve Ridge yöntemindeki alpha parametresinin en iyi değerleri bulunmak istenmektedir. Buna göre

- i. boston veri kümesini yükleyiniz
- ii. verinizi scale ediniz
- iii. Polinomsal öznelikleri oluşturun
- iv. Derece için 1, 2, 5, 10; alpha için 0.001, 0.01, 0.1, 1, 10, 100 değerlerini deneyiniz.
- v. GridSearchCV sınıfını kullanınız veridage için en iyi parametreleri belirleyiniz.
- vi. En iyi parametrelerle Ridge uygulayınız ve eğitim ve test skorunu raporlayınız.

Sonuçlarınızın ve yorumlarınızın olduğu detaylı bir rapor yazmanız gerekmektedir. Referanslarınızı rapora ekleyiniz. Tüm dosyalarınızı (rapor ve jupyter notebook) zipleyip submit ediniz.

Berna Kiraz