zenius × Kampus Merdeka INDONESIA JAYA

Hyperparameter Tuning

Day 16

Machine Learning Class

Program Studi Independen Bersertifikat
Zenius Bersama Kampus Merdeka

1. Why do we do Hyperparameter Tuning?
2. Regularization in Linear Regression
3. Hyperparameter Tuning in K-Nearest Neighbor
4. Hyperparameter Tuning in Random Forest
5. Hands-On
6. Homework Explanation

Why do we do Hyperparameter Tuning?

What are Hyperparameters?

Hyperparameter = settings that manage how a model 'learns'

Example:

- **In Random Forest, we can adjust how many decision trees that it 'creates'**
- **In Random Forest, we can adjust the maximum depth of each Decision Trees**
- **In K-Nearest Neighbor, we can adjust how many nearest neighbors to see (the number of 'k')**

Reasons to do Hyperparameter Tuning

Sometimes, tuning these settings can improve model performance.

However, please don’t fall into a mistaken mindset!

- **If a model performs poorly, doing Hyperparameter Tuning might not improve the model.**
- **Hyperparameter Tuning is our FINAL attempt to improve the accuracy of our CHOSEN model.**

Intuition

Hyperparameter Tuning is NOT:

From a 60% accuracy >>> to 90% accuracy.

Hyperparameter Tuning is:

From a 85% accuracy >>> to 88.5% accuracy.

Intuition

Asumsi tes antigen itu sekitar 400 ribu tes per hari.
Bila kita memiliki akurasi 85%, maka ada 15% = 60 ribu orang yang salah diagnosa.

Bila kita memiliki akurasi 88%, maka ‘hanya akan ada’ 48 ribu orang yang salah diagnosa.

Beda 12 ribu. 12 ribu kasus COVID yang seharusnya bisa diamankan, per hari.

Intuition

Jadi jangan meremehkan kenaikan 2-3%...

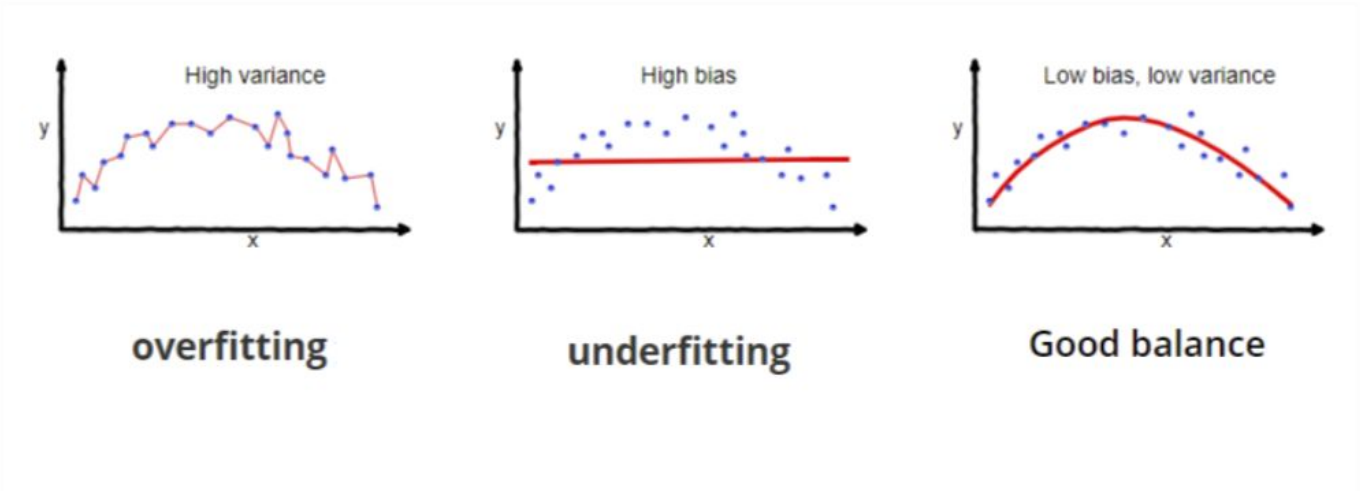
Regularization in Linear Regression

Use Case

While not essentially a ‘Hyperparameter Tuning’...

...regularization in Linear Regression attempts to avoid overfitting by giving ‘penalty’ to ‘insignificant’ features.

Use Case



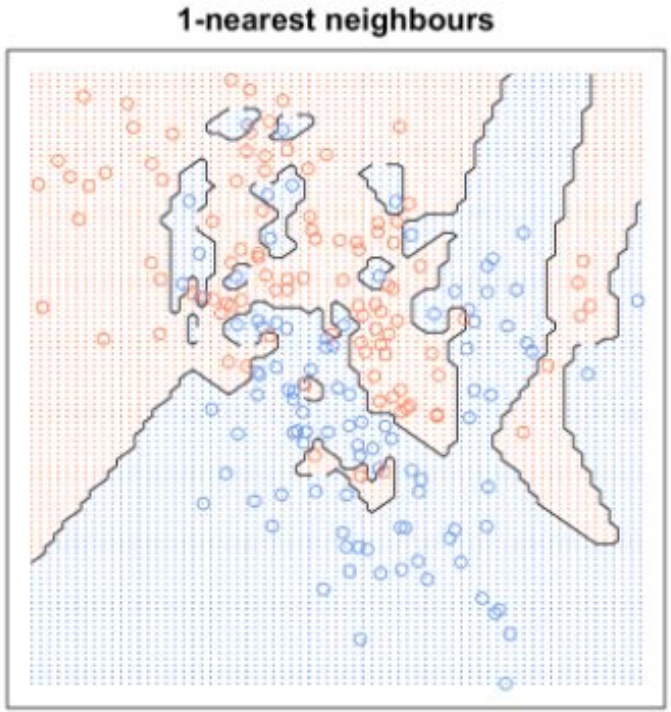
Use Case

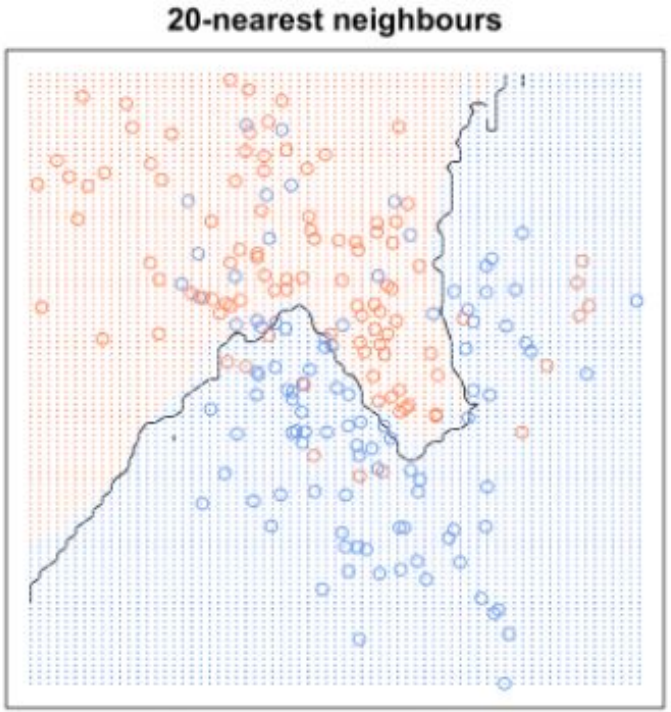
Ridge Regression: attempts to minimize regression coefficients

Lasso: if a predictor is deemed ‘insignificant’, the coefficient will be made 0

Elastic Net: combines Ridge and Lasso

K-Nearest Neighbor Tuning





Random Forest Tuning

Random Forest Hyperparameters

Guide on How to Tune Random Forest:
https://www.analyticsvidhya.com/blog/2020/03/beginners-guide-random-forest-hyperparameter-tuning/

Hyperparameter 1: Max Depth

Max Depth: How deep a decision tree will grow

- **Deeper Tree = More Complex**
- **Too Shallow = will not be able to model relationships between variables**
- **Too Deep = Overfitting**

Hyperparameter 2: n_estimators

N_estimators = number of Trees spawned

- **Too many trees = model becomes heavy, and the law of diminishing return will occur**
- **Too few trees = model tend to overfit**
- **If number of tree = 1, then it's the same as Decision Tree**

Hyperparameter 3: min_sample_split

If min_sample_split = 2, then the tree will split itself until each nodes at the bottom are completely pure.

This makes the tree grows in size and can overfit

Hands On

Hands On

Subject Matter Expert will demonstrate how to do hyperparameter tuning in the hands-on notebook. Participants are expected to follow.

Terima kasih!

Ada pertanyaan?

Assignment

Assignment

1. Selain nilai ‘k’, mana hyperparameter bawah ini yang merupakan hyperparameter pada K-Nearest Neighbors?
   a. ‘p’
   b. ‘objective’
   c. ‘probability’
   d. ‘distribution’

Assignment

2. Secara default, setiap titik di K-Nearest Neighbors memiliki nilai ‘voting’ yang sama. Namun, jika kita juga ingin memperhitungkan jarak sebagai faktor ‘voting’, hyperparameter mana yang harus diubah?
   a. ‘algorithm’
   b. ‘leaf_size’
   c. ‘weights’
   d. ‘metric’

Assignment

3. Numbers of features used merupakan sebuah Hyperparameter untuk K-Nearest Neighbors. Benar atau Salah?
   a. True
   b. False

Assignment

4. Bacalah dokumentasi package ‘lightgbm’. Apa saja bentuk dari hyperparameter tuning yang bisa dilakukan, jika kita ingin mengutamakan akurasi?
   a. Try lambda_l1 for regularization
   b. Use small num_iterations with large learning_rate
   c. Use small learning_rate with large num_iterations
   d. Use small max_bin

Assignment

5. Bacalah dokumentasi package ‘lightgbm’. Apa saja bentuk dari hyperparameter tuning yang bisa dilakukan, jika kita ingin menghindari overfitting?
   a. Use large max_bin
   b. Use feature sub-sampling
   c. Use large num_leaves
   d. Decreasing path_smooth