

Nama : Lailatul Hasanah

NIM : 222011364

Kelas : 2KS4

Kode : 13T3YVK

1. Kapan sebaiknya masing-masing teknik resampling digunakan, yaitu bootstrap, jackknife, permutation test, dan cross-validation?

Jackknife

- Jackknife digunakan Ketika ingin menghitung Standard deviation, confidence interval, dan bias. Jackknife biasa diterapkan pada skema pengambilan sampel yang kompleks, seperti pengambilan sampel multi-tahap dengan bobot pengambilan sampel yang bervariasi. Selain itu jackknife digunakan Ketika bias jauh lebih tinggi daripada varians.

Bootstrap

- Bootstrap digunakan Ketika ingin mencari Standard deviation, confidence interval, hypothesis testing, dan bias, ketika asumsi dari tes hipotesis diragukan. Selain itu, digunakan Ketika inferensi parametrik tidak mungkin atau memerlukan rumus yang sangat rumit untuk perhitungan kesalahan standar.

Permutation test

- Permutation Test digunakan untuk Hypothesis testing, Ketika asumsi awal tidak diperlukan (kecuali dapat dipertukarkan), dan distribusi sampling dari statistik pengujian di bawah H_0 dihitung dengan menggunakan permutasi data.

Cross validation

- Cross Validation digunakan untuk Model validation yang bertujuan untuk memperoleh hasil akurasi yang maksimal.

2. Apa perbedaan utama antara bootstrap parametrik dan non parametrik?

Perbedaan utama antara bootstrap parametrik dan non parametrik adalah untuk bootstrap parametrik memiliki asumsi informasi tentang distribusi populasinya, sedangkan bootstrap non parametrik tidak membuat asumsi tentang distribusi populasinya. Jika asumsi tentang distribusi populasi benar maka bootstrap parametrik akan memberi hasil yang lebih baik daripada bootstrap nonparametrik. Jika tidak benar, maka bootstrap nonparametrik akan bekerja lebih baik.

3. Mengapa cross-validation bermanfaat untuk mencegah dan mengatasi overfitting?

Karena cross validation digunakan untuk memperkirakan kinerja prediktif model pada data yang tidak terlihat. Cross validation hanya melakukan ini dengan lebih kuat, dengan mengulangi eksperimen beberapa kali, menggunakan semua bagian berbeda dari set pelatihan sebagai set validasi. Ini memberikan indikasi yang

lebih akurat tentang seberapa baik model menggeneralisasi data yang tidak terlihat. Dengan kata lain, validasi silang tidak mencegah overfitting itu sendiri, tetapi dapat membantu dalam mengidentifikasi kasus overfitting.

4. Pada kondisi seperti apa sebaiknya masing-masing teknik pembagian data digunakan, yaitu metode split, LOOV, dan k-fold cross validation?
 - Split digunakan Ketika ingin mencari sekumpulan variabel yang memberikan error rate testing terendah dengan kumpulan data yang besar.
 - LOOV digunakan Ketika ingin mendapatkan bias dan MSE yang lebih rendah nilainya dibandingkan metode split. Hanya saja karena metode ini mahal tidak cocok untuk kumpulan data yang besar.
 - K-fold digunakan Ketika ingin mencapai tujuan mendekati metode LOOV (MSE dan bias rendah), meskipun dengan kumpulan data yang dimiliki besar. Metode ini merupakan gabungan dari metode LOOV dan Split.