

Support Vector Machine

Procedure 方法

Support Vector Machine adalah sebuah algoritma machine learning yang bekerja dengan membangun sebuah hyperplane yang memaksimalkan jarak antar titik kelas.

1. Membangun hyperplane

Hyperplane dicari sehingga memisahkan data dari kelas yang berbeda secara maksimal. Titik paling dekat dengan hyperplane disebut support vectors, dan jaraknya disebut margin.

2. Mengaplikasikan Kernel ke Data

Kernel adalah transformasi yang dilakukan ke data. Untuk kernel non-linear, transformasi kernel digunakan untuk memetakan data ke ruang dimensi lebih tinggi sehingga lebih separable jika ingin digambar hyperplane dibandingkan representasi linear. Kernel non-linear yang di-implementasikan disini adalah Radial Basis Function, yang melakukan pemetaan dengan fungsi Gaussian.

3. Optimasi Hyperplane dengan Stochastic Gradient Descent

Optimasi disini adalah memaksimalkan margin, implementasi from scratch disini menggunakan Stochastic Gradient Descent dari hinge loss.

4. Buat SVM untuk tiap pasangan kelas

Karena SVM hanya bisa bekerja dengan binary classification, untuk multiclass classification disini diimplementasikan one vs one classifier, sehingga dibuat classifier untuk tiap pasangan kelas.

5. Prediksi

Prediksi dilakukan dengan hasil voting semua classifier. Per-classifier prediksi dilakukan dengan menentukan di sisi mana dari hyperplane prediksi tersebut jatuh.

VS Sklearn

Hasil di notebook DoE menghasilkan nilai F1 yang lebih tinggi. Hal tersebut mungkin dikarenakan perbandingan dilakukan dengan SGDClassifier hinge untuk mencocokkan implementasi SGD, walaupun SVM biasanya yang digunakan sklearn menggunakan solver SMO yang lebih baik.

Potential Improvements

- Menggunakan solver SMO agar lebih cepat dan lebih baik
- Hyperparameter tuning dari parameter