

Mašīnmācīšanās: Klasifikators Kurš?

Problēmas apraksts

Ivans apmācīja septiņus dažādus ML modeļus attēlu klasifikācijas datu kopai: Decision Tree, K-Nearest Neighbors, Logistic Regression, Gaussian Process Classifier, Multi-Layer Perceptron, Support Vector Machine un Random Forest.

Viņš atstāja savu datoru neatslēgtu pusdienu laikā. Alise, nolēmusi pamulķoties, sajaukāja prognožu failu nosaukumus un pielaboja viņa apmācības hiperparametrus.

"Alise! Tagad nevaru pateikt, kuras prognozes nāk no kura modeļa!" Ivans protestēja.

"Tu esi ML eksperts," Alise pasmaidīja. "Vai Tev nevajadzētu spēt to noskaidrot pēc to uzvedības?"

Ivans saprata, ka viņai ir taisnība. Tas patiesībā bija interesants forenzs uzdevums — un ideāls olimpiādes uzdevums.

Uzdevums

Jums ir dota attēlu datu kopa `x_train.csv` kopā ar to apzīmējumiem (labels) `y_train.csv`, uz kurām tika apmācīti 6 modeļi.

Modeļi veica prognozes uz testa kopas (test set) `x_test.csv` (un `y_test.csv`), un prognozes ir saglabātas failos `preds_0.csv`, `preds_1.csv`, ..., `preds_6.csv`. Izmantotie modeļi:

- Decision Tree, DT
- K-Nearest Neighbors, KNN
- Logistic Regression, LR
- Multi-Layer Perceptron, MLP
- Support Vector Machine, SVM
- Random Forest, RF

Padomi

- Izmantojiet sklearn, lai apmācītu modeļus un veiktu prognozes.
- Pārbaudiet, kuri modeļi sniedz visvairāk līdzīgas prognozes ar dotajiem prognožu failiem.
- Izmantojiet PCA no sklearn, lai samazinātu dimensionalitāti LR un MLP.
- Ja neesat pazīstami ar dažiem modeļiem vai ar sklearn kopumā, iesakam sākt ar mūsu galveno ceļvedi par to, kā sākt ar ML. Tas būs noderīgs visiem ML uzdevumiem.

Vērtēšana

Vērtēts ar precizitāti modeļu identificēšanā (piem., 2 modeļi atrasti pareizi = 2/6).