

Vytěžování dat, cvičení 9: k-Nearest Neighbour Classifier

Michael Anděl, Miroslav Čepek



Evropský sociální fond
Praha & EU: Investujeme do vaší budoucnosti

Fakulta elektrotechnická, ČVUT

1. V prostředí Matlab implementujte k -NN klasifikátor a otestujte jej na datasetu „Breast Cancer“ 3-fold krosvalidace¹. Alternativně můžete validovat mode pomocí testovací sady (cca 1-3), kterou z data na začátku vydělíte.
2. Vzorky s chybějícími hodnotami (označené "?") můžete výjimečně zahodit.
3. Vyčíslete trénovací i testovací chyby klasifikátoru pro hodnoty k od 1 do velikosti datasetu. Oba průběhy vynesete do grafu v závislosti na k .
4. Průběh grafu interpretujte s přihlédnutím k principu fungování algoritmu k -NN.

Zdrojové kódy odevzdejte do upload systému samostatně vedle PDF protokolu.

¹viz. Wikipedie