

Horní odhady klasifikační chyby

Návod k desátému cvičení

ZS 2011/2012

Na přednáškách jste se seznámili s metodou pro určování pravděpodobnostních horních odhadů chyby klasifikátorů. V této úloze si vyzkoušíte aplikovat tuto teorii na problém učení se konjunkcí.

Pro teoretický odhad chyby budeme vycházet z následující nerovnosti, která platí s pravděpodobností $1 - \delta$ (při splnění předpokladů uvedených na přednášce):

$$e(f) \leq e(S, f) + \sqrt{\frac{1}{2|S|} \ln \frac{2|\mathcal{F}|}{\delta}}$$

Zde $e(f)$ označuje skutečnou chybu, \mathcal{F} je konceptový prostor a $e(S, f)$ odhad chyby na trénovací množině S .

Zadání

Vášim úkolem bude vypočítat horní odhad chyby pro konjunkce délky nejvýše k . Tento odhad zobrazíte do jednoho grafu spolu s několika empiricky změřenými křivkami učení. Toto provedete pro několik různých k .

Křivky učení konstruuje pomocí funkce `learningCurve`, kterou si můžete stáhnout ze stránek cvičení (`10_matlab.zip`)¹. Funkce `conj_learningCurve` přebírá na svém vstupu šest parametrů.

```
function lc = conj_learningCurve(learningFunction, k, ...  
    train, test, train_labels, test_labels)
```

První parametr je handle na funkci pro učení se konjunkcí, druhý je maximální délka hledaných konjunkcí, třetí a čtvrtý parametr jsou trénovací, resp. testovací data ve formě matic (kde 1 je TRUE a -1 je FALSE) a konečně pátý a šestý argument jsou vektory se zařazením příkladů buď mezi pozitivní (1), nebo negativní (-1).

K vygenerování trénovacích a testovacích dat použijte připravený kód v souboru `cviceni_10.m`. Konjunkce budete hledat pomocí algoritmu `conj_bb`, který jste

¹Tato funkce se liší od stejnojmenné funkce používané v minulém cvičení tím, že křivka učení, kterou tato funkce vrací, není průměrem několika křivek.

používali i v minulém cvičení. Jednu křivku učení vygenerujete například následujícím kódem:

```
lc = conj_learningCurve(@conj_bb, size(train,2),...  
    train, test, train_labels, test_labels);
```