



FAKULTA
APLIKOVANÝCH VĚD
ZÁPADOČESKÉ
UNIVERZITY
V PLZNI

KATEDRA
KYBERNETIKY



Metody Počítačového Vidění

8. cvičení

Obsah cvičení

1. Klasifikace
2. Recall, Precision, F1 score

1 Klasifikace

Ve strojovém učení a statistice je klasifikace problém identifikace, ke kterému ze souboru kategorií patří nová pozorování a to na základě trénovací množiny dat. Příkladem by mohl být antispamový filtr, který na základě jména klasifikuje do složek „je spam“ nebo „není spam“.

V terminologii strojového učení je klasifikace za instanci učení s učitelem, tj učení, kde je k dispozici trénovací sada správně označených pozorování. Podobně, klasifikaci, kde není informace od učitele se říká shlukování a zahrnuje seskupování dat do kategorií na základě jisté míry přirozené podobnosti nebo vzdálenosti.

Často mají jednotlivá pozorování sadu měřitelných vlastností. Tyto vlastnosti mohou představovat jednotlivé kategorie například typ krevní skupiny, pořadové číslo („malé“, „střední“, „velké“).

Algoritmus, který provádí klasifikaci a to zejména v konkrétním provedení je známý jako klasifikátor. Termín „klasifikátor“ někdy také odkazuje na matematické funkce, realizované klasifikačním algoritmem, který mapuje vstupní data do kategorií.

Terminologie skrze vědecké oblasti je velice pestrá. Ve statistice, kde se klasifikace často provádí binární regresí nebo obdobným postupem se pozorované proměnné nazývají „nezávislé proměnné“ a predikované kategorie jsou známé jako výsledky, které je možné považovat za závislé proměnné. Ve strojovém učení jsou pozorování známá jako instance a nezávislé proměnné jsou nazývány příznaky, které se sdružují do příznakových vektorů. Dále jsou kategorie nazývány třídami.

2 Recall, Precision, F1 score

		Actual class	
		1	0
Predicted class	1	True Positive	False Positive
	0	False Negative	True Negative

Precision

Precision is the probability that a (randomly selected) retrieved document is relevant.

$$\text{Precision} = \frac{\text{True positives}}{\#\text{predicted as positive}} = \frac{\text{True positives}}{\text{True positives} + \text{False positives}} \quad (1)$$

Recall

Recall is the probability that a (randomly selected) relevant document is retrieved in a search.

$$\text{Recall} = \frac{\text{True positives}}{\#\text{actual positives}} = \frac{\text{True positives}}{\text{True positives} + \text{False negatives}} \quad (2)$$

F1 score

A measure that combines precision and recall is the harmonic mean of precision and recall.

$$\text{F1 score} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \quad (3)$$