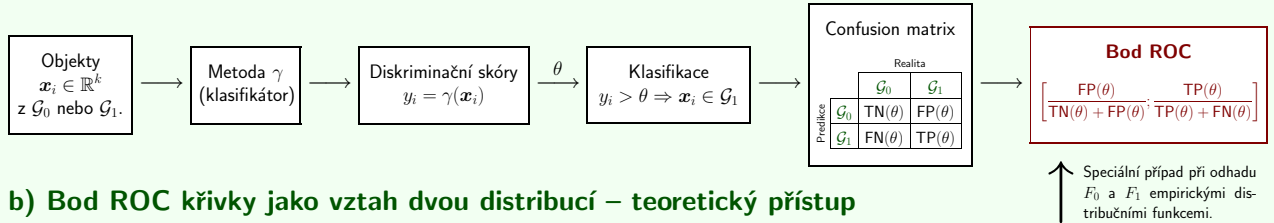


Poznámky k ROC křivkám

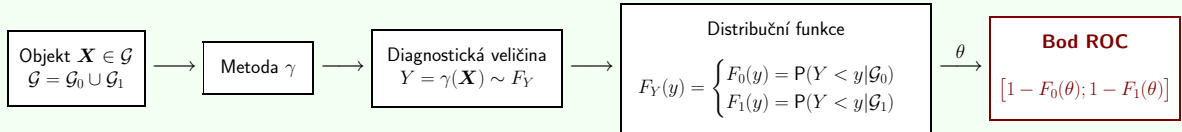
Martin Betinec & Luboš Prchal
betinec@matfyz.cz & prchal@karlin.mff.cuni.cz

Tento poster je věnován vybraným otázkám, jež se týkají statistické analýzy ROC křivek, resp. užití ROC křivek např. v klasifikaci. Nejprve však náš dvojí pohled na samotnou ROC křivku:

a) Bod ROC křivky jako míra úspěšnosti klasifikace – aplikovaný přístup



b) Bod ROC křivky jako vztah dvou distribucí – teoretický přístup



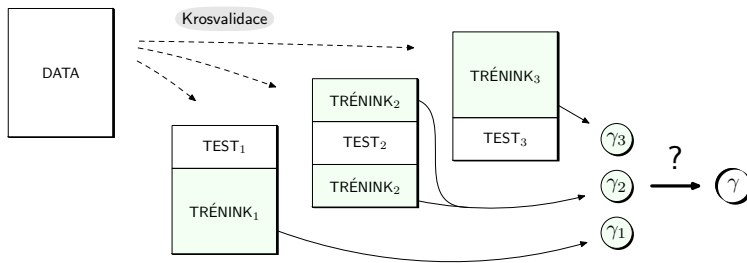
ROC křivka je množina ROC bodů pro všechny hodnoty $\theta \in \mathbb{R}$ – charakterizuje chování klasifikátoru.

#1 Odhad ROC křivky

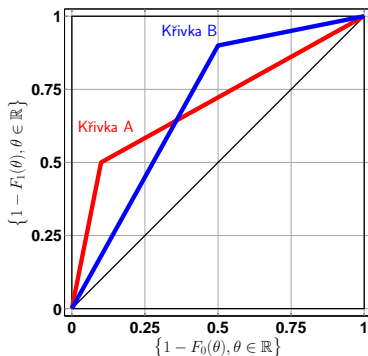
Většina složitějších klasifikátorů (metod γ) musí projít fází učení z trénovacích dat. Testovací soubor následně slouží k ověření kvality nastavení klasifikátoru.

- Každé trénování (na části dat) však typicky vyústí v jiné nastavení klasifikátoru – v jiný klasifikátor – $\gamma_1, \gamma_2, \gamma_3$.
- Z testovací fáze tudíž pro jednotlivé γ_i dostaneme rozdílné ROC křivky.

! Lze na základě získaných ROC křivek hodnotit globální chování uvažované metody ?



U křivek rovnocenných dle kvantit. kritérií (AUC, vzdálenost od bodu $[0, 1]$) záleží na preferencích.



#2 Shoda ROC křivek

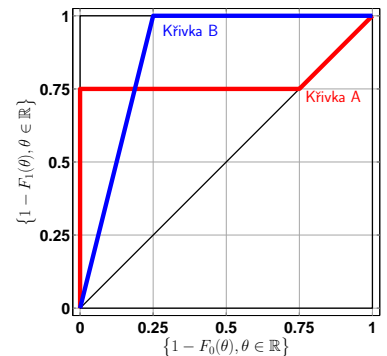
Pracujeme-li s různými metodami, jejichž kvalitu měříme ROC křivkami, vyvstává základní otázka: Jsou ROC křivky odpovídající použitým metodám shodné?

Některá užívaná kritéria:

- Vzdálenost od bodu $[0, 1]$
- AUC – Area Under Curve
- Preference uživatele – konzervativní \times liberální metoda

! Lze kombinací uvedených kritérií rozhodnout o shodě, resp. významné odlišnosti, dvou ROC křivek ?

Považujeme-li TP rate $\leq 75\%$ za dostatečný, lze metodu A preferovat, přestože $AUC_B > AUC_A$.



#3 Výběr prediktorů

Nechť je analyzovaný objekt popsán k prediktory. Mějme pravidlo (související s metodou γ), jak postupně odstraňovat nejméně významné prediktory.

Předpokládejme, že vyřazením nevýznamného prediktoru nedojde k významné změně ROC křivky, a naopak při vyřazení podstatného prediktoru se ROC křivka změní.

! Může analýza ROC křivek přispět k optimálnímu výběru prediktorů vzhledem ke klasifikaci ?

