Classifying probabilistic distributions

S.E. Volodin

MIPT

The dataset

1. Probability distributions $\{f_k(x)\colon \mathbb{R} \to \mathbb{R}\}_{k=1}^{K}$
2. Picking one of them $k \in \overline{1, K}$
3. Picking points $[a, b] \subseteq \mathrm{Dom} f_k$
4. Dividing $[a, b]$ into 101 parts: $[a, b] = \bigsqcup\limits_{s=0}^{100} \Delta_s,$

$$|\Delta_0| = |\Delta_{100}| = \frac{\Delta_s}{2}$$

5. Sampling $t_z \sim f_k$
6. Counting numbers in Δ_s: $x_s^i = \sum\limits_z [t_z \in \Delta_s]$

$$\Rightarrow \text{Obtained object } \{x_s\} \in \mathfrak{D}$$

The dataset

Answers: $y_i = [k = \hat{k}]$ for a fixed $\hat{k}$.

$$f_{\hat{k}} = ax^{a-1}\left[x \in [0,1]\right]$$

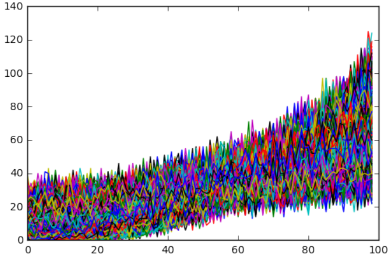


Рис.: 1. Training set, $y_i = 0$

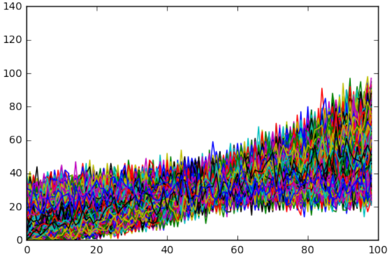
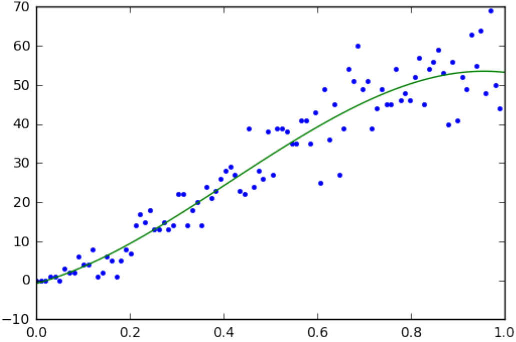


Рис.: 2. Training set, $y_i = 1$

The solution

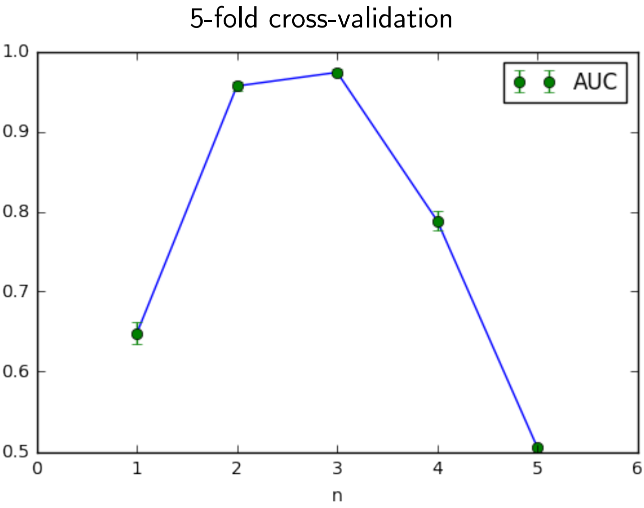
Idea: dimension reduction

1. Fitting a polynom $\sum |P_n(s) - x_s|^2 \to \min$



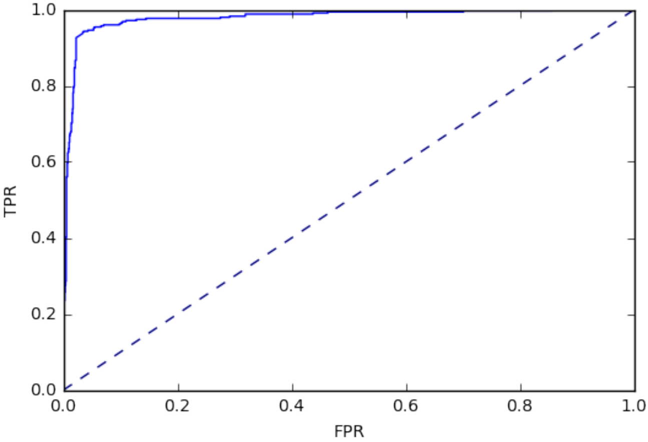
2. Using coefficients $\mathrm{coeff}(P_n)$ as features of $\{x_s^i\} \in \mathfrak{D}$

Choosing degree



Answer: $n^* = 3$

Result



Result

#	Δ2d	Team Name	Score
1	—	[illegible]	0.99450
2	—	[illegible]	0.99412
3	—	[illegible]	0.98403
4	—	**SergeyVolodin**	**0.97144**

The end

Thank you!
Questions?