

# Классификация вероятностных распределений

С.Е. Володин

МФТИ

- 1 Вероятностные распределения  $\{f_k(x): \mathbb{R} \rightarrow \mathbb{R}\}_{k=1}^K$
- 2 Выбирается  $k \in \overline{1, K}$
- 3 Выбираются  $[a, b] \subseteq \mathbb{R}$
- 4  $[a, b]$  разбивается на 101 часть:  $[a, b] = \bigsqcup_{s=0}^{100} \Delta_s$ ,

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

- 5 Генерируются  $t_z \sim f_k$
- 6 Вычисляется количество объектов в  $\Delta_s$ :  $x_s^i = \sum_z [t_z \in \Delta_s]$

$\Rightarrow$  Получен объект выборки  $\{x_s\} \in \mathfrak{D}$

Ответы:  $y_i = [k = \hat{k}]$  для фиксированного  $\hat{k}$ .

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

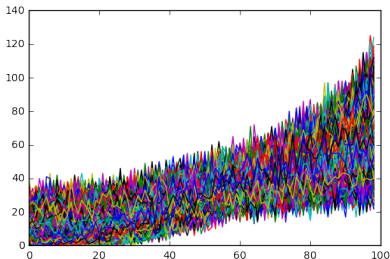


Рис.: 1. Обучающая выборка,  
 $y_i = 0$

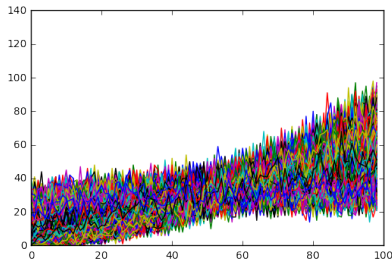
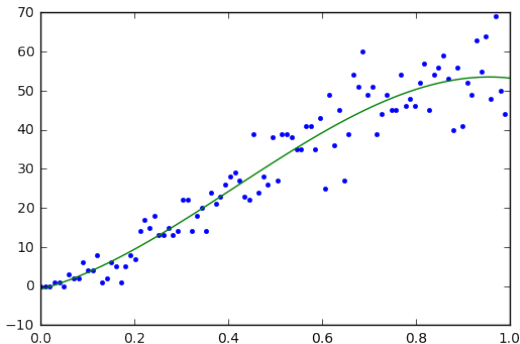


Рис.: 2. Обучающая выборка,  
 $y_i = 1$

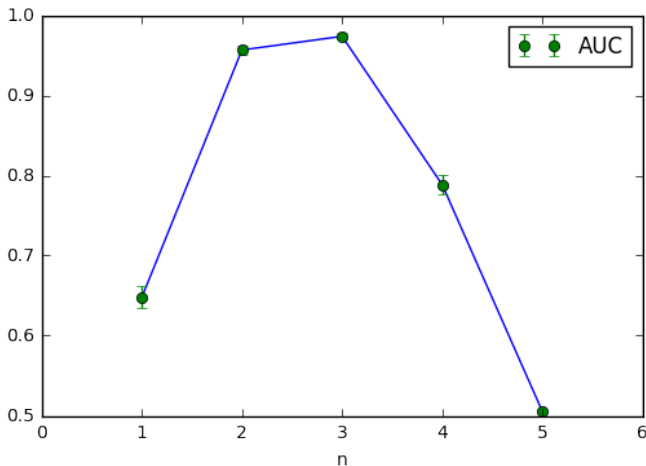
Идея: снижение размерности

- ① Аппроксимация многочленом  $\sum |P_n(s) - x_s|^2 \rightarrow \min$

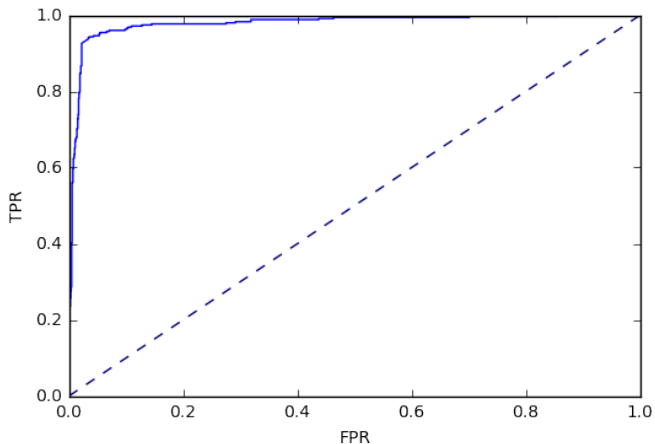


- ② Использование коэффициентов  $\text{coeff}(P_n)$  как признаков объекта  $\{x_s^i\} \in \mathfrak{D}$

## 5-кратная кросс-валидация



Ответ:  $n^* = 3$



#	$\Delta 2d$	Team Name	Score ?
1	—		0.99450
2	—		0.99412
3	—		0.98403
4	—	<b>SergeyVolodin</b>	<b>0.97144</b>

Спасибо за внимание.  
Вопросы?