

Решение задачи соревнования по курсу «Машинное обучение»

Денисов Д.М.

Механико-математический факультет МГУ имени М.В. Ломоносова

Восстановление пропусков

Данные разнородны, заменять на одно значение смысла мало
 \Rightarrow используем KNN

Параметры: 5 соседей, веса пропорциональны расстояниям.

Настройка классификатора

Выбираем классификатор KNN и проводим 5-fold кросс-валидацию на случайной десятой части тренировочной выборки по следующей сетке гиперпараметров:

```
algorithm      : [kd_tree, ball_tree]  
n_neighbors : [1, 3, 5, 7, 9]  
weights       : [distance]
```

Получаем:

```
algorithm      : kd_tree  
n_neighbors : 9  
weights       : distance
```

Количество соседей получилось равным максимальному
⇒ может быть и больше...

Настройка классификатора

Меняем сетку гиперпараметров, увеличивая максимальное число соседей:

```
algorithm      : [kd_tree, ball_tree]  
n_neighbors    : [1, ..., 20]  
weights        : [distance]
```

Получаем:

```
algorithm      : kd_tree  
n_neighbors    : 13  
weights        : distance
```

Действительно, 13 соседей подходят больше.

Обучение

Обучаем на десятой части тренировочной выборки
 \Rightarrow ассигасу $\sim 90,5\%$.

Обучаем на всей тренировочной выборке
 \Rightarrow ассигасу $\sim 95,4\%$.