

# Листок 02. Прогнозирование

Н.В. Артамонов

9 июня 2024 г.

## Содержание

<b>1</b>	<b>Линейная регрессия</b>	<b>1</b>
<b>2</b>	<b>k-NN</b>	<b>6</b>
<b>3</b>	<b>Валидация моделей</b>	<b>11</b>

## 1 Линейная регрессия

#1. Для набора данных `sleep75` рассмотрим линейную регрессию

`sleep` на `totwrk`, `age`, `south`, `male`.

1. Подгоните модель и выведите коэффициенты подогнанной модели
2. Рассмотрим трёх людей с характеристиками

index	totwrk	age	south	male
0	2160	32	1	0
1	1720	24	0	1
2	2390	44	0	1

вычислите прогноз `sleep`

3. На обучающей выборке вычислите метрики подгонки:  $R^2$ , MSE, MAE, MAPE, RMSE

#2. Для набора данных `sleep75` рассмотрим линейную регрессию

**sleep** на **totwrk**, **age**, **south**, **male**, **smsa**, **yngkid**, **marr**.

1. Подгоните модель и выведите коэффициенты подогнанной модели
2. Рассмотрим трёх людей с характеристиками

index	totwrk	age	south	male	smsa	yngkid	marr
0	2150	37	0	1	1	0	1
1	1950	28	1	1	0	1	0
2	2240	26	0	0	1	0	0

вычислите прогноз **sleep**

3. На обучающей выборке вычислите метрики подгонки:  $R^2$ , MSE, MAE, MAPE, RMSE

#3. Для набора данных `wage2` рассмотрим линейную регрессию

**wage** на **age**, **IQ**, **south**, **married**, **urban**.

1. Подгоните модель и выведите коэффициенты подогнанной модели
2. Рассмотрим трёх людей с характеристиками

index	age	IQ	south	married	urban
0	36	105	1	1	1
1	29	123	0	1	0
2	25	112	1	0	1

вычислите прогноз **wage**

3. На обучающей выборке вычислите метрики подгонки:  $R^2$ , MSE, MAE, MAPE, RMSE

#4. Для набора данных `wage2` рассмотрим линейную регрессию

**log(wage)** на **age**, **IQ**, **south**, **married**, **urban**.

1. Подгоните модель и выведите коэффициенты подогнанной модели

2. Рассмотрим трёх людей с характеристиками

index	age	IQ	south	married	urban
0	36	105	1	1	1
1	29	123	0	1	0
2	25	112	1	0	1

вычислите прогноз **wage**

3. На обучающей выборке вычислите метрики подгонки:  $R^2$ , MSE, MAE, MAPE, RMSE

#5. Для набора данных **wage1** рассмотрим линейную регрессию

**wage** на **exper**, **female**, **married**, **smsa**.

1. Подгоните модель и выведите коэффициенты подогнанной модели
2. Рассмотрим трёх людей с характеристиками

index	exper	female	married	smsa
0	5	1	1	1
1	26	0	0	1
2	38	1	1	0

вычислите прогноз **wage**

3. На обучающей выборке вычислите метрики подгонки:  $R^2$ , MSE, MAE, MAPE, RMSE

#6. Для набора данных **wage1** рассмотрим линейную регрессию

**log(wage)** на **exper**, **female**, **married**, **smsa**.

1. Подгоните модель и выведите коэффициенты подогнанной модели
2. Рассмотрим трёх людей с характеристиками

index	exper	female	married	smsa
0	5	1	1	1
1	26	0	0	1
2	38	1	1	0

вычислите прогноз **wage**

3. На обучающей выборке вычислите метрики подгонки:  $R^2$ , MSE, MAE, MAPE, RMSE

#7. Для набора данных **Labour** рассмотрим линейную регрессию

**output на capital, labour.**

1. Подгоните модель и выведите коэффициенты подогнанной модели
2. Рассмотрим три фирмы с характеристиками

index	capital	labour
0	2.970	85
1	10.450	60
2	3.850	105

вычислите прогноз **output**

3. На обучающей выборке вычислите метрики подгонки:  $R^2$ , MSE, MAE, MAPE, RMSE

#8. Для набора данных **Labour** рассмотрим линейную регрессию

**log(output) на log(capital), log(labour).**

1. Подгоните модель и выведите коэффициенты подогнанной модели
2. Рассмотрим три фирмы с характеристиками

index	capital	labour
0	2.970	85
1	10.450	60
2	3.850	105

вычислите прогноз **output**

3. На обучающей выборке вычислите метрики подгонки:  $R^2$ , MSE, MAE, MAPE, RMSE

#9. Для набора данных Labour рассмотрим линейную регрессию

**output на capital, labour, wage.**

1. Подгоните модель и выведите коэффициенты подогнанной модели
2. Рассмотрим три фирмы с характеристиками

index	capital	labour	wage
0	2.970	85	36.98
1	10.450	60	33.82
2	3.850	105	40.23

вычислите прогноз **output**

3. На обучающей выборке вычислите метрики подгонки:  $R^2$ , MSE, MAE, MAPE, RMSE

#10. Для набора данных Labour рассмотрим линейную регрессию

**$\log(\text{output})$  на  $\log(\text{capital})$ ,  $\log(\text{labour})$ ,  $\log(\text{wage})$ .**

1. Подгоните модель и выведите коэффициенты подогнанной модели
2. Рассмотрим три фирмы с характеристиками

index	capital	labour	wage
0	2.970	85	36.98
1	10.450	60	33.82
2	3.850	105	40.23

вычислите прогноз **output** для каждой

3. На обучающей выборке вычислите метрики подгонки:  $R^2$ , MSE, MAE, MAPE, RMSE

## 2 k-NN

#1. Для набора данных **sleep75** рассмотрим задачу прогнозирования для переменных

зависимая/target	объясняющая/предикторы/features
sleep	totwrk, age, south, male

1. подгоните на исходном датасете модель k-NN с параметрами

№	$k$	веса
1	5	uniform
2	5	distant
3	10	uniform
4	10	distant

2. Рассмотрим трёх людей с характеристиками

index	totwrk	age	south	male
0	2160	32	1	0
1	1720	24	0	1
2	2390	44	0	1

вычислите прогноз **sleep** по каждой модели

#2. Для набора данных **sleep75** рассмотрим задачу прогнозирования для переменных

зависимая/target	объясняющая/предикторы/features
sleep	totwrk, age, south, male, smsa, yngkid, marr

1. подгоните на исходном датасете модель k-NN с параметрами

№	$k$	веса
1	5	uniform
2	5	distant
3	10	uniform
4	10	distant

2. Рассмотрим трёх людей с характеристиками

index	totwrk	age	south	male	smsa	yngkid	marr
0	2150	37	0	1	1	0	1
1	1950	28	1	1	0	1	0
2	2240	26	0	0	1	0	0

вычислите прогноз **sleep** по каждой модели

#3. Для набора данных **wage2** рассмотрим задачу прогнозирования для переменных

зависимая/target	объясняющая/предикторы/features
wage	age, IQ, south, married, urban

1. подгоните на исходном датасете модель k-NN с параметрами

№	$k$	веса
1	5	uniform
2	5	distant
3	10	uniform
4	10	distant

2. Рассмотрим трёх людей с характеристиками

index	age	IQ	south	married	urban
0	36	105	1	1	1
1	29	123	0	1	0
2	25	112	1	0	1

вычислите прогноз **wage** по каждой модели

#4. Для набора данных **wage2** рассмотрим задачу прогнозирования для переменных

зависимая/target	объясняющая/предикторы/features
log(wage)	age, IQ, south, married, urban

1. подгоните на исходном датасете модель k-NN с параметрами

№	$k$	веса
1	5	uniform
2	5	distant
3	10	uniform
4	10	distant

2. Рассмотрим трёх людей с характеристиками

index	age	IQ	south	married	urban
0	36	105	1	1	1
1	29	123	0	1	0
2	25	112	1	0	1

вычислите прогноз **wage** по каждой модели

#5. Для набора данных **wage1** рассмотрим задачу прогнозирования для переменных

зависимая/target	объясняющая/предикторы/features
wage	exper, female, married, smsa

1. подгоните на исходном датасете модель k-NN с параметрами

№	$k$	веса
1	5	uniform
2	5	distant
3	10	uniform
4	10	distant

2. Рассмотрим трёх людей с характеристиками

index	exper	female	married	smsa
0	5	1	1	1
1	26	0	0	1
2	38	1	1	0

вычислите прогноз **wage** по каждой модели

#6. Для набора данных **wage1** рассмотрим задачу прогнозирования для переменных



зависимая/target	объясняющая/предикторы/features
$\log(\text{wage})$	exper, female, married, smsa

1. подгоните на исходном датасете модель k-NN с параметрами

№	$k$	веса
1	5	uniform
2	5	distant
3	10	uniform
4	10	distant

2. Рассмотрим трёх людей с характеристиками

index	exper	female	married	smsa
0	5	1	1	1
1	26	0	0	1
2	38	1	1	0

вычислите прогноз **wage** по каждой модели

#7. Для набора данных **Labour** рассмотрим задачу прогнозирования для переменных

зависимая/target	объясняющая/предикторы/features
output	capital, labour

1. подгоните на исходном датасете модель k-NN с параметрами

№	$k$	веса
1	5	uniform
2	5	distant
3	10	uniform
4	10	distant

2. Рассмотрим трёх людей с характеристиками

index	capital	labour
0	2.970	85
1	10.450	60
2	3.850	105

вычислите прогноз **output** по каждой модели

#8. Для набора данных **Labour** рассмотрим задачу прогнозирования для переменных

зависимая/target	объясняющая/предикторы/features
log(output)	log(capital), log(labour)

1. подгоните на исходном датасете модель k-NN с параметрами

№	$k$	веса
1	5	uniform
2	5	distant
3	10	uniform
4	10	distant

2. Рассмотрим трёх людей с характеристиками

index	capital	labour
0	2.970	85
1	10.450	60
2	3.850	105

вычислите прогноз **output** по каждой модели

#9. Для набора данных **Labour** рассмотрим задачу прогнозирования для переменных

зависимая/target	объясняющая/предикторы/features
output	capital, labour, wage

1. подгоните на исходном датасете модель k-NN с параметрами

№	$k$	веса
1	5	uniform
2	5	distant
3	10	uniform
4	10	distant

2. Рассмотрим трёх людей с характеристиками

index	capital	labour	wage
0	2.970	85	36.98
1	10.450	60	33.82
2	3.850	105	40.23

вычислите прогноз **output** по каждой модели

#10. Для набора данных **Labour** рассмотрим задачу прогнозирования для переменных

зависимая/target	объясняющая/предикторы/features
log(output)	log(capital), log(labour), log(wage)

1. подгоните на исходном датасете модель k-NN с параметрами

№	$k$	веса
1	5	uniform
2	5	distant
3	10	uniform
4	10	distant

2. Рассмотрим трёх людей с характеристиками

index	capital	labour	wage
0	2.970	85	36.98
1	10.450	60	33.82
2	3.850	105	40.23

вычислите прогноз **output** по каждой модели

### 3 Валидация моделей

#1. Набор данных **sleep75** разбейте на обучающую и тестовую часть в соотношении 80:20.

Рассмотрим задачу прогнозирования для переменных

зависимая/target	объясняющая/предикторы/features
sleep	totwrk, age, south, male

и следующие модели

№	Модель
1	линейная регрессия
2	k-NN с $k = 5$ , веса 'uniform'
3	k-NN с $k = 5$ , веса 'distant'
4	k-NN с $k = 10$ , веса 'uniform'
5	k-NN с $k = 10$ , веса 'distant'

Проведите валидацию моделей относительно метрик  $R^2$ , MSE, MAE, MAPE. Какая модель предпочтительней?

#2. Набор данных `sleep75` разбейте на обучающую и тестовую часть в соотношении 80:20.

Рассмотрим задачу прогнозирования для переменных

зависимая/target	объясняющая/предикторы/features
sleep	totwrk, age, south, male, smsa, yngkid, marr

и следующие модели

№	Модель
1	линейная регрессия
2	k-NN с $k = 5$ , веса 'uniform'
3	k-NN с $k = 5$ , веса 'distant'
4	k-NN с $k = 10$ , веса 'uniform'
5	k-NN с $k = 10$ , веса 'distant'

Проведите валидацию моделей относительно метрик  $R^2$ , MSE, MAE, MAPE. Какая модель предпочтительней?

#3. Набор данных `wage2` разбейте на обучающую и тестовую часть в соотношении 80:20.

Рассмотрим задачу прогнозирования для переменных

зависимая/target	объясняющая/предикторы/features
wage	age, IQ, south, married, urban

и следующие модели

№	Модель
1	линейная регрессия
2	k-NN с $k = 5$ , веса 'uniform'
3	k-NN с $k = 5$ , веса 'distant'
4	k-NN с $k = 10$ , веса 'uniform'
5	k-NN с $k = 10$ , веса 'distant'

Проведите валидацию моделей относительно метрик  $R^2$ , MSE, MAE, MAPE. Какая модель предпочтительней?

#4. Набор данных **wage2** разбейте на обучающую и тестовую часть в соотношении 80:20.

Рассмотрим задачу прогнозирования для переменных

зависимая/target	объясняющая/предикторы/features
$\log(\text{wage})$	age, IQ, south, married, urban

и следующие модели

№	Модель
1	линейная регрессия
2	k-NN с $k = 5$ , веса 'uniform'
3	k-NN с $k = 5$ , веса 'distant'
4	k-NN с $k = 10$ , веса 'uniform'
5	k-NN с $k = 10$ , веса 'distant'

Проведите валидацию моделей относительно метрик  $R^2$ , MSE, MAE, MAPE. Какая модель предпочтительней?

#5. Набор данных **wage1** разбейте на обучающую и тестовую часть в соотношении 80:20.

Рассмотрим задачу прогнозирования для переменных

зависимая/target	объясняющая/предикторы/features
wage	exper, female, married, smsa

и следующие модели

№	Модель
1	линейная регрессия
2	k-NN с $k = 5$ , веса 'uniform'
3	k-NN с $k = 5$ , веса 'distant'
4	k-NN с $k = 10$ , веса 'uniform'
5	k-NN с $k = 10$ , веса 'distant'

Проведите валидацию моделей относительно метрик  $R^2$ , MSE, MAE, MAPE. Какая модель предпочтительней?

#6. Набор данных **wage1** разбейте на обучающую и тестовую часть в соотношении 80:20.

Рассмотрим задачу прогнозирования для переменных

зависимая/target	объясняющая/предикторы/features
log(wage)	exper, female, married, smsa

и следующие модели

№	Модель
1	линейная регрессия
2	k-NN с $k = 5$ , веса 'uniform'
3	k-NN с $k = 5$ , веса 'distant'
4	k-NN с $k = 10$ , веса 'uniform'
5	k-NN с $k = 10$ , веса 'distant'

Проведите валидацию моделей относительно метрик  $R^2$ , MSE, MAE, MAPE. Какая модель предпочтительней?

#7. Набор данных **Labour** разбейте на обучающую и тестовую часть в соотношении 80:20.

Рассмотрим задачу прогнозирования для переменных

зависимая/target	объясняющая/предикторы/features
output	capital, labour, wage

и следующие модели

№	Модель
1	линейная регрессия
2	k-NN с $k = 5$ , веса 'uniform'
3	k-NN с $k = 5$ , веса 'distant'
4	k-NN с $k = 10$ , веса 'uniform'
5	k-NN с $k = 10$ , веса 'distant'

Проведите валидацию моделей относительно метрик  $R^2$ , MSE, MAE, MAPE. Какая модель предпочтительней?

#8. Набор данных **Labour** разбейте на обучающую и тестовую часть в соотношении 80:20.

Рассмотрим задачу прогнозирования для переменных

зависимая/target	объясняющая/предикторы/features
log(output)	log(capital), log(labour), log(wage)

и следующие модели

№	Модель
1	линейная регрессия
2	k-NN с $k = 5$ , веса 'uniform'
3	k-NN с $k = 5$ , веса 'distant'
4	k-NN с $k = 10$ , веса 'uniform'
5	k-NN с $k = 10$ , веса 'distant'

Проведите валидацию моделей относительно метрик  $R^2$ , MSE, MAE, MAPE. Какая модель предпочтительней?