

## 2 Azure ML Studio

### 2.1 Модель линейной регрессии

Для создания модели машинного обучения с помощью Azure ML Studio мы рекомендуем придерживаться следующих этапов:

1. Импортировать данные в формате CSV.
2. Выбрать колонки, которые будут использоваться в обучении модели.
3. Выбрать метод (блок) регрессии или классификации.
4. Обучить модель, выбрав столбец, соответствующий отклику.

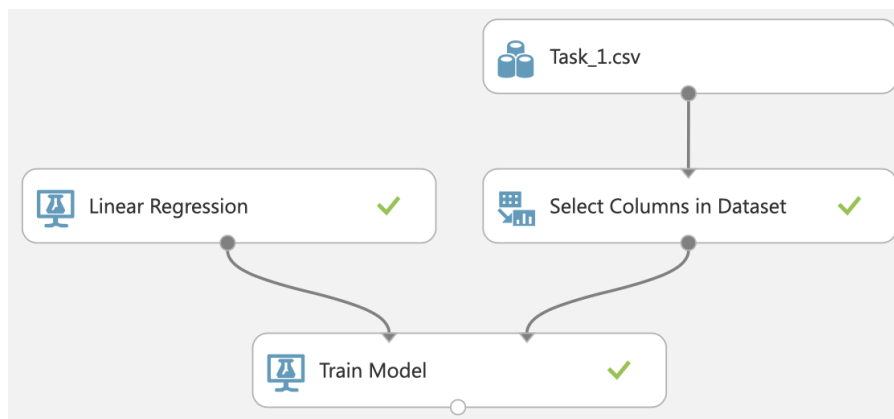


Рис. 1: Схема модели линейной регрессии.

### 2.2 Обучение модели

Для обучения модели линейной регрессии используется блок **Linear Regression** из раздела **Machine Learning**. Единственный параметр, который нас будет интересовать – **Random number seed**. Этот параметр позволяет получать одинаковые результаты значений коэффициентов и параметров модели, предполагающих их численное нахождение

Блок **Train Model** отвечает за обучение модели. На вход подаются данные и выбранный метод машинного обучения. В параметрах данного блока необходимо выбрать столбец данных, который мы будем предсказывать (отклик).

После запуска модели, полученные значения коэффициентов уравнения линейной регрессии можно посмотреть в параметрах блока **Train Model**, пункт **Visualize**. Параметр **Bias** (смещение) соответствует коэффициенту  $\theta_0$ , а название колонки в данных –  $\theta_1$ .

## Feature Weights

Feature	Weight
Bias	4.03905
Количество человек в очереди	2.11433

Рис. 2: Коэффициенты уравнения линейной регрессии.

## 2.3 Задача предсказания

Для решения задачи предсказания необходимы данные. В качестве данных могут выступать либо данные в формате CSV, из которых надо отобрать только предикторы, либо это могут быть введенные вручную значения с помощью блока **Enter Data Manually**. Обратите внимание, первая строка должна содержать название столбцов таблицы, а далее построчно следуют данные, для которых будем строить прогноз:

Количество человек в очереди

27

28

25

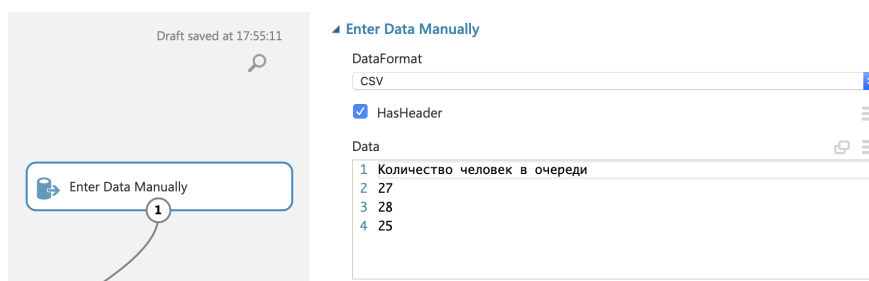


Рис. 3: Ввод данных вручную (один столбец).

Если столбцов несколько, то следует разделить запятой названия колонок и соответствующие им значения.

Data
1 competitorname,chocolate,fruity,caramel,peanutyalmondy,nougat,crispedricewafer,hard,bar,pluribus,sugarperc
2 100 Grand,1,0,1,0,0,1,0,1,0,0.73199999,0.86000001,66.971725,1
3 3 Musketeers,1,0,0,0,1,0,0,1,0,0.60399997,0.51099998,67.602936,1
4 One dime,0,0,0,0,0,0,0,0,0,0.011,0.116,32.261086,0
5 One quarter,0,0,0,0,0,0,0,0,0,0.011,0.51099998,46.116505,0
6 Air Heads,0,1,0,0,0,0,0,0,0,0.90600002,0.51099998,52.341465,1

Рис. 4: Ввод данных вручную (несколько столбцов).

Блоки, необходимо соединить с помощью **Score Model**.

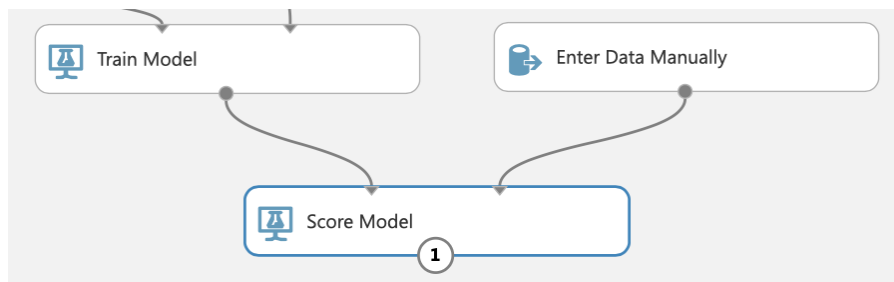


Рис. 5: Блоки необходимые для выполнения прогноза.

Далее эксперимент следует запустить, используя **Run**, а результаты прогноза можно посмотреть в пункте **Visualize** блока **Score Model**


	Количество человек в очереди	Scored Labels
view as  	 	 
	27	61.126006
	28	63.240338
	25	56.897343

Рис. 6: Результаты прогноза.

### 3 Оценка модели

За оценку модели отвечает блок **Evaluate Model**, подключаемый к **Score Model**. После обучения модели и запуска доступны метрики, рассмотренные в лекции:

- Среднее квадратическое отклонение остатков **RSE** (Relative Squared Error);
- $R^2$  статистика (Coefficient of Determination).

## Metrics

Mean Absolute Error	5.8158
Root Mean Squared Error	9.003265
Relative Absolute Error	0.510926
Relative Squared Error	0.457258
Coefficient of Determination	0.542742

Рис. 7: Оценка модели.

## 4 Модель многомерной линейной регрессии

В случае, если интересующий нас показатель зависит более, чем от одной переменной, следует обучить модель многомерной линейной регрессии. Для этого достаточно выбрать из набора данных любое число столбцов, на которых будет происходить обучение модели, аналогично линейной.

Select columns

BY NAME  
WITH RULES

AVAILABLE COLUMNS

All Types search columns

competitorname  
chocolate  
fruity  
caramel  
peanutyalmondy  
nougat  
crispedricewafer  
hard  
bar  
pluribus  
Y

SELECTED COLUMNS

All Types search columns

sugarpercent  
pricepercent  
winpercent

11 columns available

3 columns selected

Рис. 8: Выбор данных для построения модели многомерной линейной регрессии.