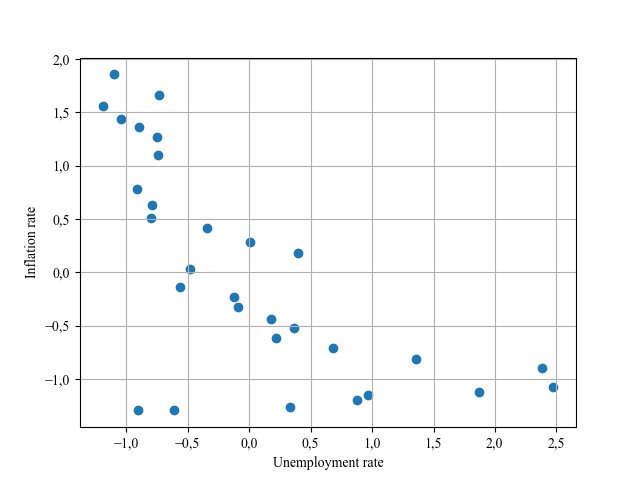
Кривая Филлипса для страны Российская Федерация

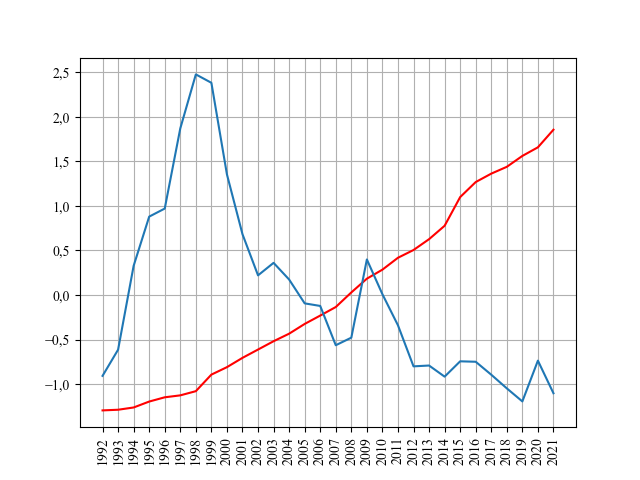
Получим данные о динамике безработицы(Inflation rate (%)) и динамике безработицы(Unemployment rate (%)) из статистики Мирового Банка о стране Российская Федерация.

|  |  |  |
| --- | --- | --- |
| date | Inflation rate | Unemployment rate |
| 1992 | 0.053595784106737 | 5.181 |
| 1993 | 0.52215463296409 | 5.883 |
| 1994 | 2.12894263118429 | 8.131 |
| 1995 | 6.33177914447035 | 9.449 |
| 1996 | 9.35533109407782 | 9.665 |
| 1997 | 10.7363023210616 | 11.813 |
| 1998 | 13.7087206048547 | 13.261 |
| 1999 | 25.4634679089422 | 13.036 |
| 2000 | 30.7595536540075 | 10.581 |
| 2001 | 37.3657852105703 | 8.978 |
| 2002 | 43.2653684455772 | 7.875 |
| 2003 | 49.1768425137193 | 8.21 |
| 2004 | 54.5315199244747 | 7.763 |
| 2005 | 61.4490089758499 | 7.124 |
| 2006 | 67.3903013768284 | 7.055 |
| 2007 | 73.4603471089943 | 6.002 |
| 2008 | 83.8261661028973 | 6.205 |
| 2009 | 93.5896759401721 | 8.301 |
| 2010 | 100 | 7.369 |
| 2011 | 108.440464859326 | 6.536 |
| 2012 | 113.943539767608 | 5.436 |
| 2013 | 121.638956306017 | 5.458 |
| 2014 | 131.155272814079 | 5.16 |
| 2015 | 151.529464141176 | 5.571 |
| 2016 | 162.200847296653 | 5.559 |
| 2017 | 168.175238863746 | 5.212 |
| 2018 | 173.015822116403 | 4.846 |
| 2019 | 180.750263654162 | 4.496 |
| 2020 | 186.862621885623 | 5.589 |
| 2021 | 199.372063343799 | 4.715 |

Стандартизируем полученнные данные и выведем получившеееся распределение точек



Проведем анализ изменения инфляции и безработицы в стране.



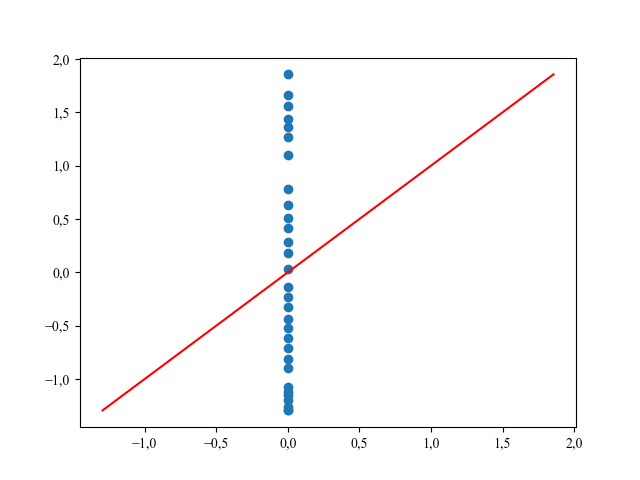
Как видно из графиков, за рассматриваемый период инфляция достигла максимального значения в 2021 году, а безработица в 1998 году, а минимумы в 1992 и 2019 годах соответственно для инфляции и безработицы.

Построим модель линейной регрессии на получившихся данных  
  
 OLS Regression Results   
==============================================================================  
Dep. Variable: y R-squared: 0.476  
Model: OLS Adj. R-squared: 0.457  
Method: Least Squares F-statistic: 25.41  
Date: Sun, 03 Nov 2024 Prob (F-statistic): 2.48e-05  
Time: 21:07:53 Log-Likelihood: -32.881  
No. Observations: 30 AIC: 69.76  
Df Residuals: 28 BIC: 72.57  
Df Model: 1   
Covariance Type: nonrobust   
==============================================================================  
 coef std err t P>|t| [0.025 0.975]  
------------------------------------------------------------------------------  
const -1.214e-16 0.137 -8.87e-16 1.000 -0.280 0.280  
x1 -0.6897 0.137 -5.041 0.000 -0.970 -0.409  
==============================================================================  
Omnibus: 4.568 Durbin-Watson: 0.201  
Prob(Omnibus): 0.102 Jarque-Bera (JB): 3.102  
Skew: -0.749 Prob(JB): 0.212  
Kurtosis: 3.487 Cond. No. 1.00  
==============================================================================  
  
Notes:  
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.Как видно, качество модели мало

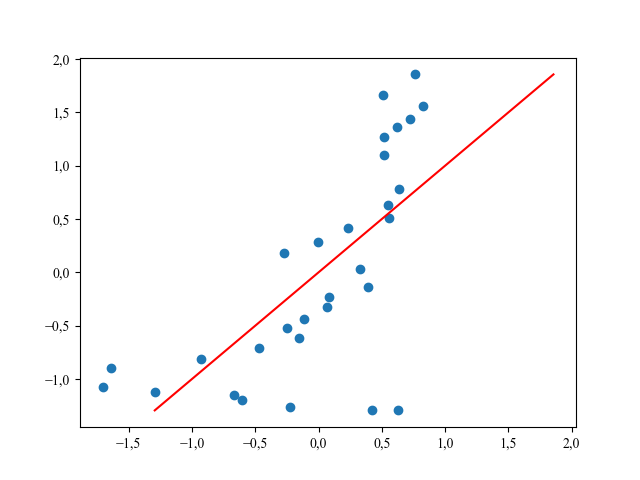
Построим модель гиперболической функции:  
  
 OLS Regression Results   
==============================================================================  
Dep. Variable: y R-squared: 0.000  
Model: OLS Adj. R-squared: -0.035  
Method: Least Squares F-statistic: 0.009688  
Date: Sun, 03 Nov 2024 Prob (F-statistic): 0.922  
Time: 21:07:53 Log-Likelihood: -42.563  
No. Observations: 30 AIC: 89.13  
Df Residuals: 28 BIC: 91.93  
Df Model: 1   
Covariance Type: nonrobust   
==============================================================================  
 coef std err t P>|t| [0.025 0.975]  
------------------------------------------------------------------------------  
x1 0.0009 0.009 0.098 0.922 -0.017 0.019  
const -0.0030 0.191 -0.015 0.988 -0.395 0.389  
==============================================================================  
Omnibus: 5.862 Durbin-Watson: 0.016  
Prob(Omnibus): 0.053 Jarque-Bera (JB): 2.365  
Skew: 0.350 Prob(JB): 0.307  
Kurtosis: 1.816 Cond. No. 21.9  
==============================================================================  
  
Notes:  
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.Как видно, качество модели также мало

Попробуем совершить полиномиальные преобразования

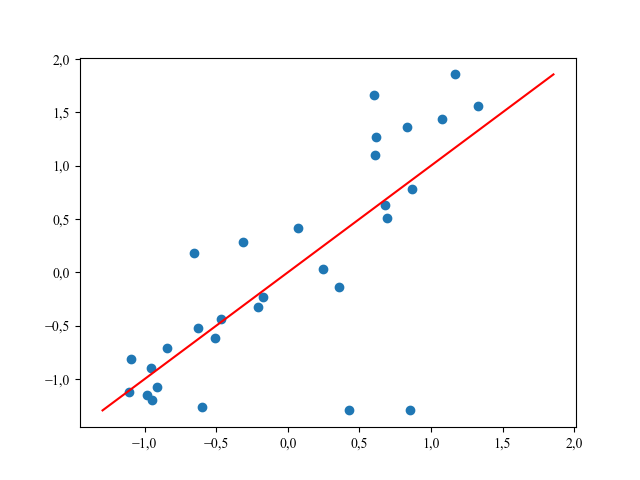
Степень полинома = 0,  
R2 значение модели = -2.220446049250313e-16



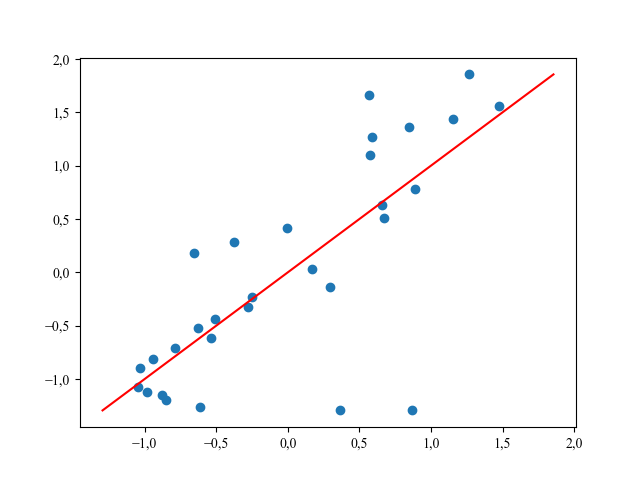
Степень полинома = 1,  
R2 значение модели = 0.4757456570052696



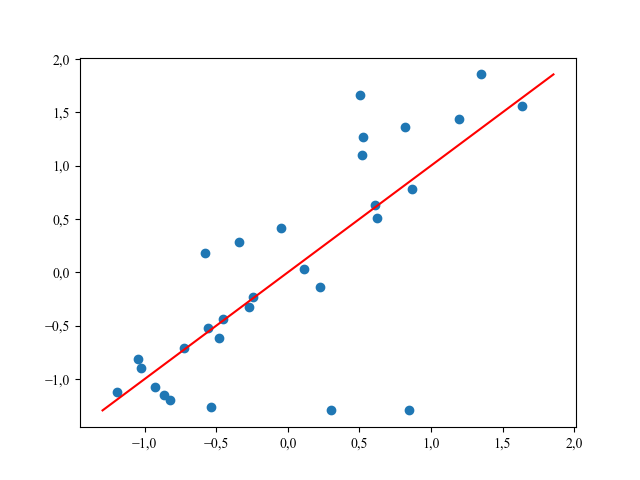
Степень полинома = 2,  
R2 значение модели = 0.5834159252740272



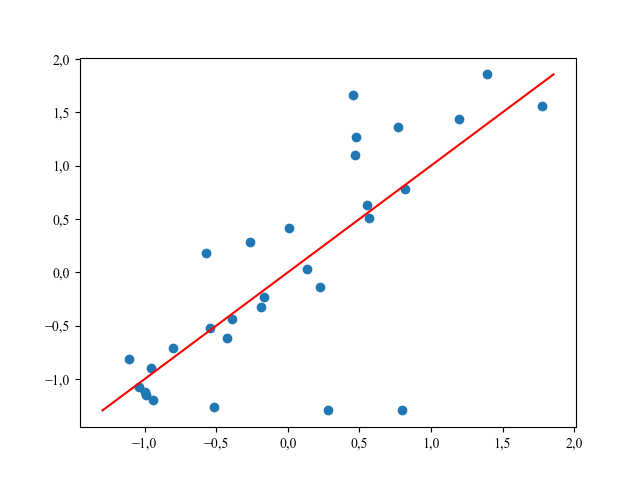
Степень полинома = 3,  
R2 значение модели = 0.5888873104428667



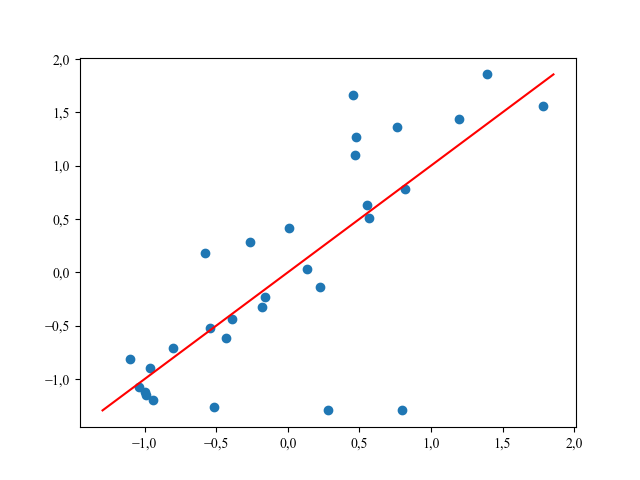
Степень полинома = 4,  
R2 значение модели = 0.5942658580403255



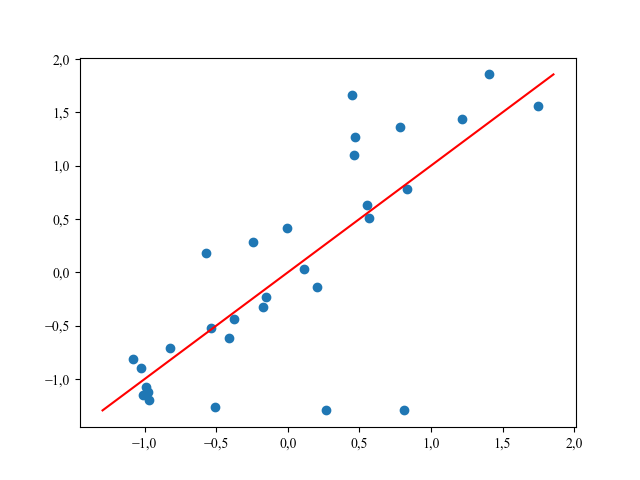
Степень полинома = 5,  
R2 значение модели = 0.5998113461035516



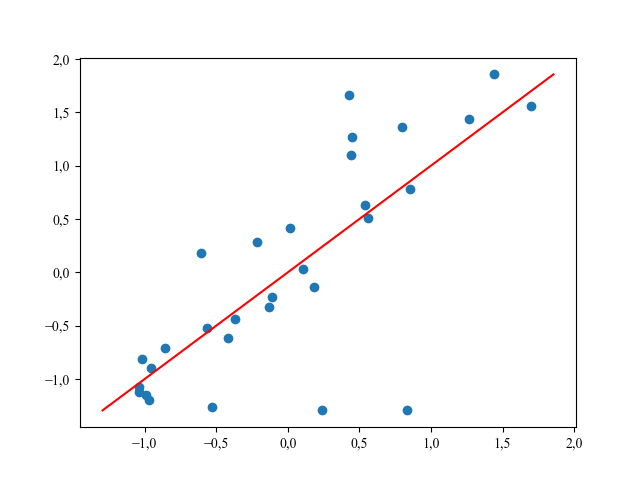
Степень полинома = 6,  
R2 значение модели = 0.5998249369279183



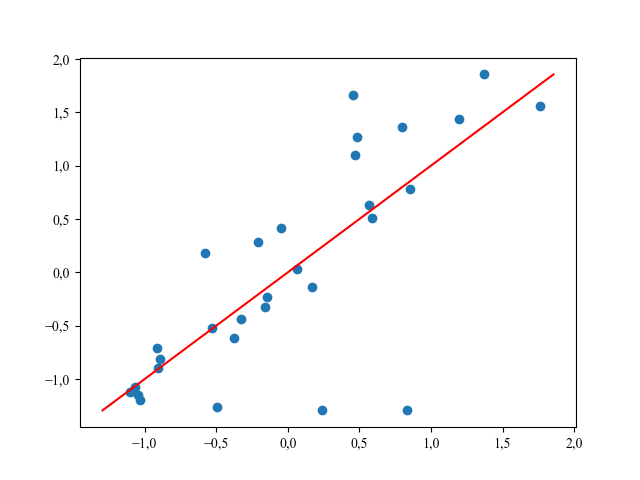
Степень полинома = 7,  
R2 значение модели = 0.6003265203730428



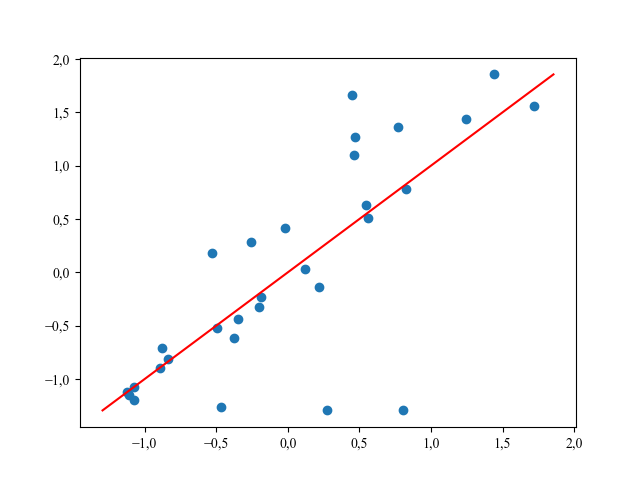
Степень полинома = 8,  
R2 значение модели = 0.6014559756196767



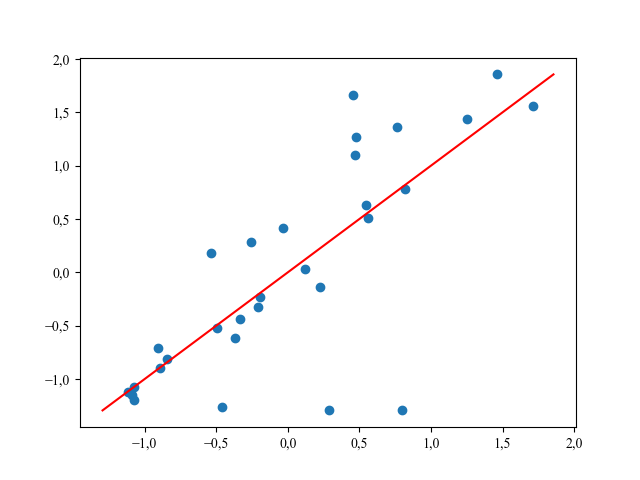
Степень полинома = 9,  
R2 значение модели = 0.6036525148575846



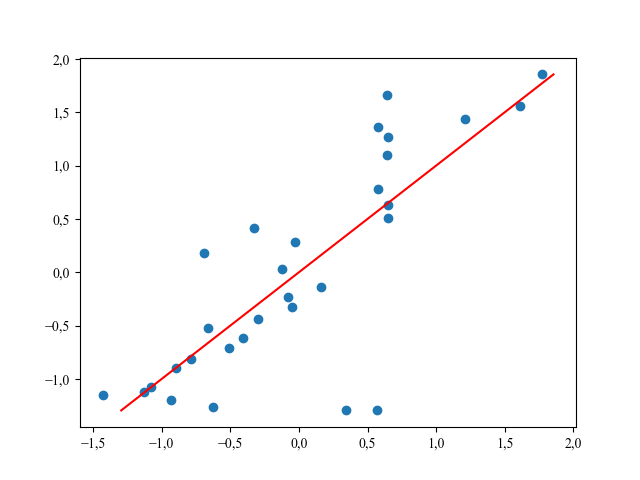
Степень полинома = 10,  
R2 значение модели = 0.6050405912338555



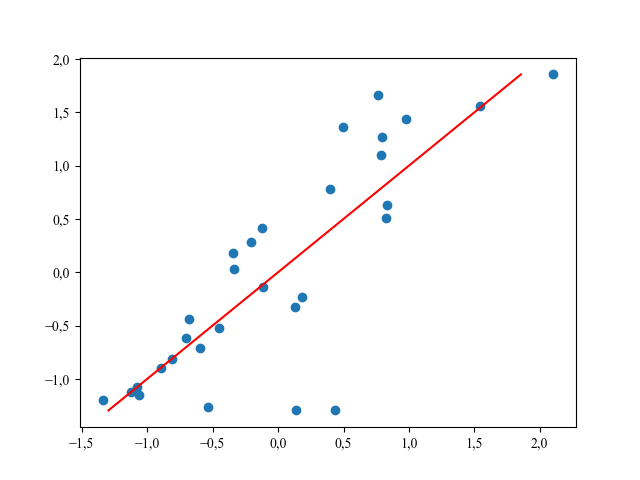
Степень полинома = 11,  
R2 значение модели = 0.6051288384650452



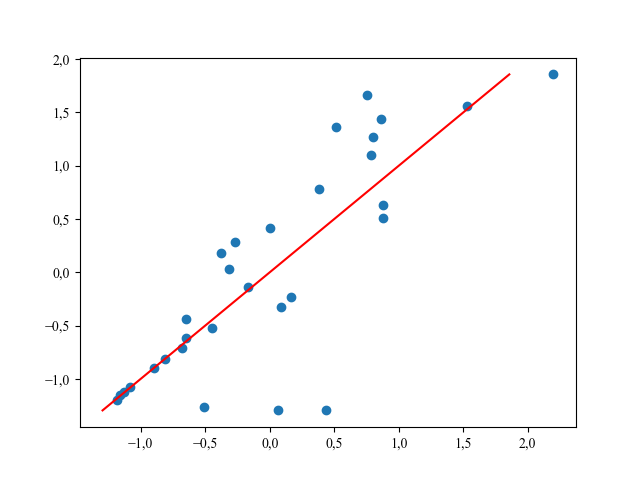
Степень полинома = 12,  
R2 значение модели = 0.6383053617154058



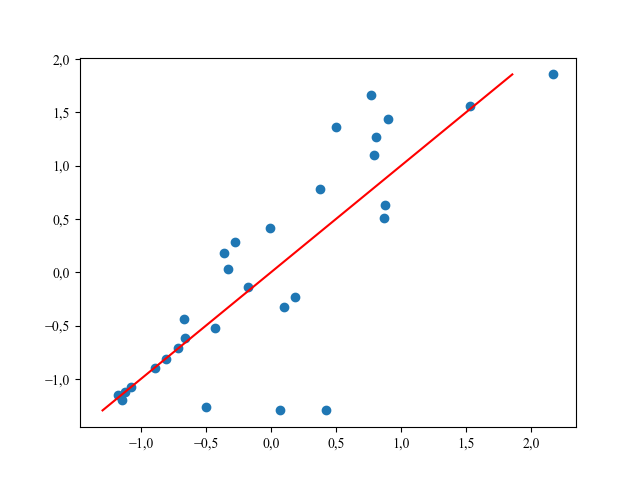
Степень полинома = 13,  
R2 значение модели = 0.6854201528826397



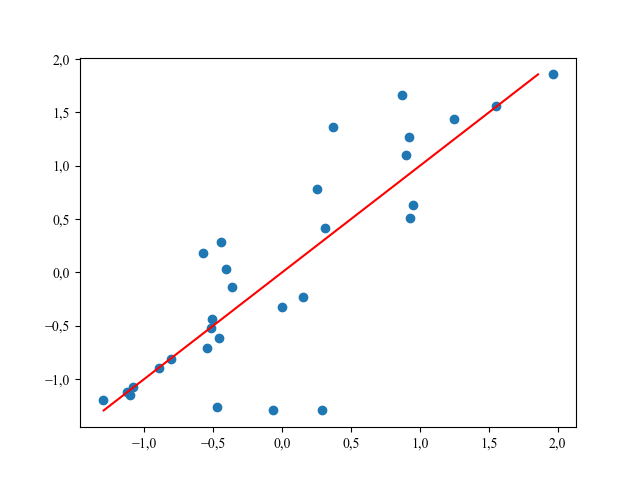
Степень полинома = 14,  
R2 значение модели = 0.6888239894153205



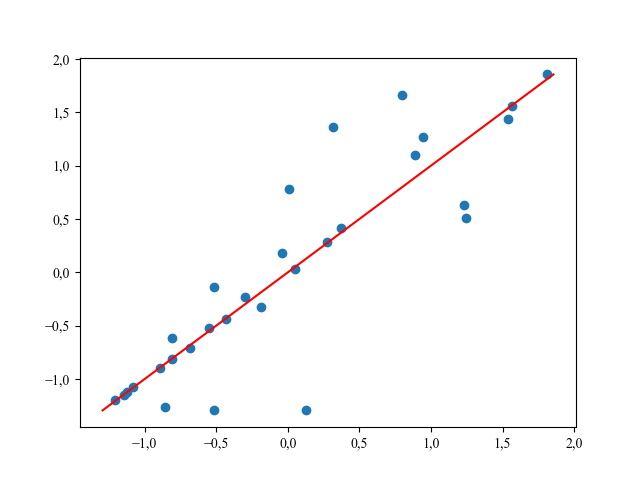
Степень полинома = 15,  
R2 значение модели = 0.6890768959443111



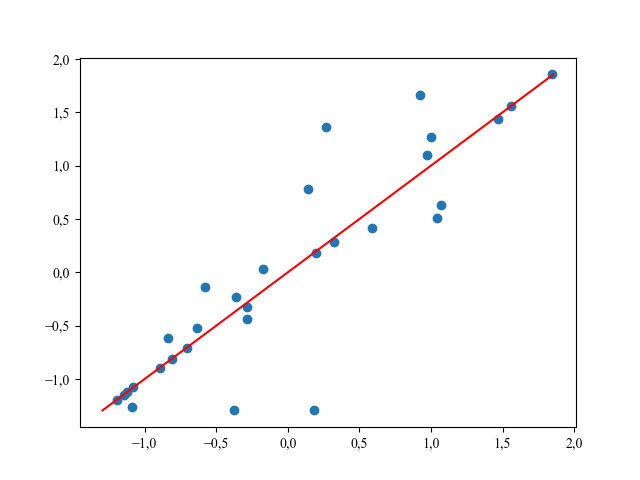
Степень полинома = 16,  
R2 значение модели = 0.7100904449584963



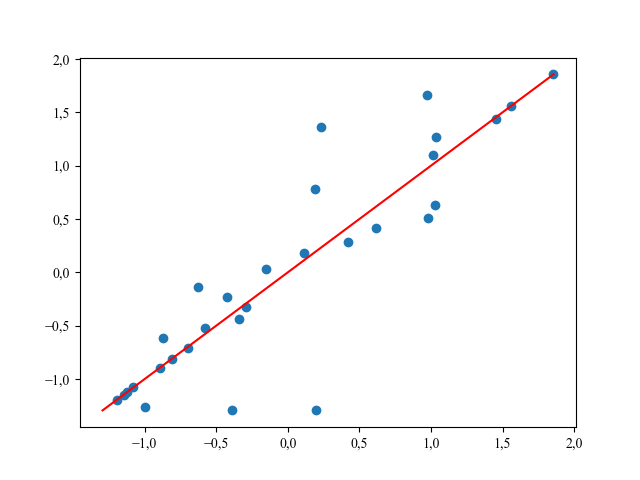
Степень полинома = 17,  
R2 значение модели = 0.7820968097786463



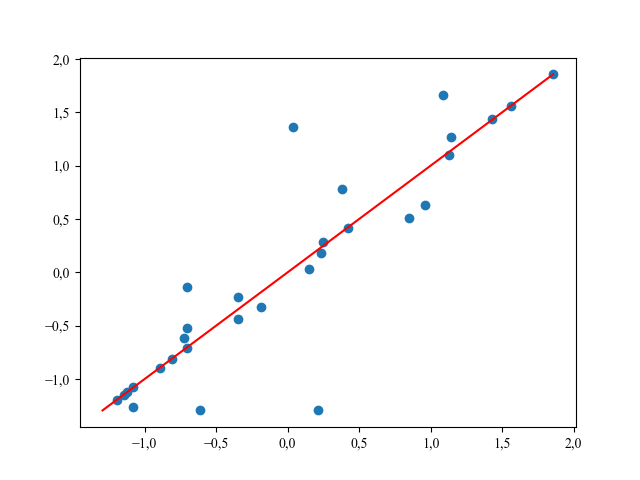
Степень полинома = 18,  
R2 значение модели = 0.7952558621819248



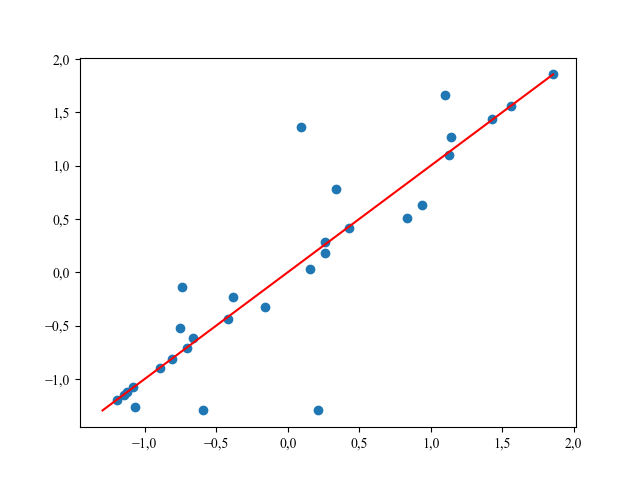
Степень полинома = 19,  
R2 значение модели = 0.7972251674655648



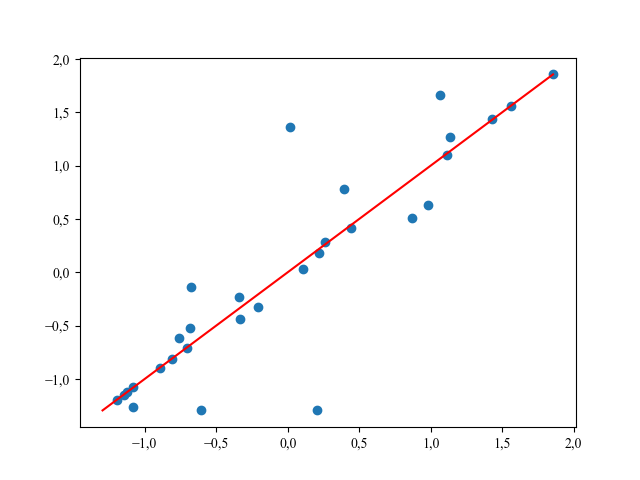
Степень полинома = 20,  
R2 значение модели = 0.8111654264888933



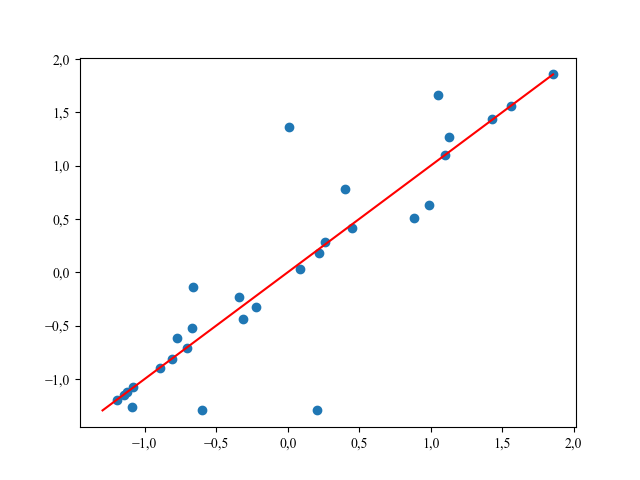
Степень полинома = 21,  
R2 значение модели = 0.8118870776612848



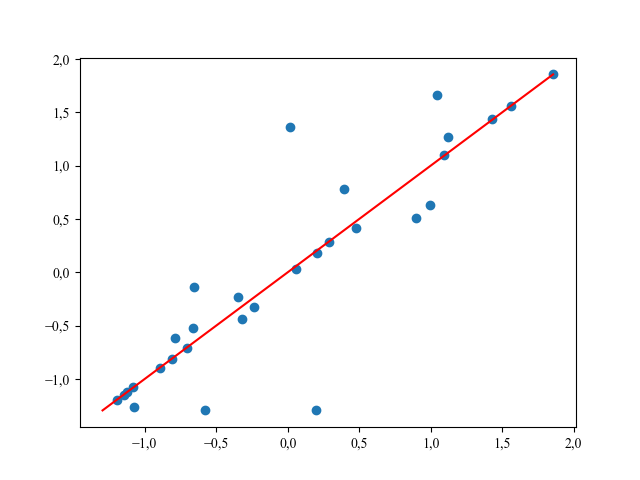
Степень полинома = 22,  
R2 значение модели = 0.8091466041814731



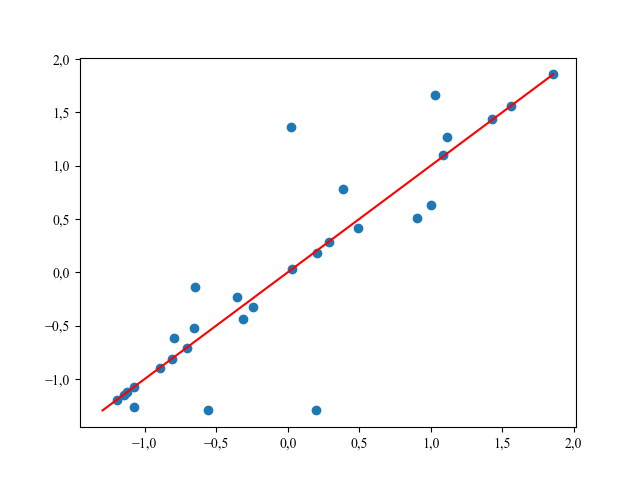
Степень полинома = 23,  
R2 значение модели = 0.8081909946979868



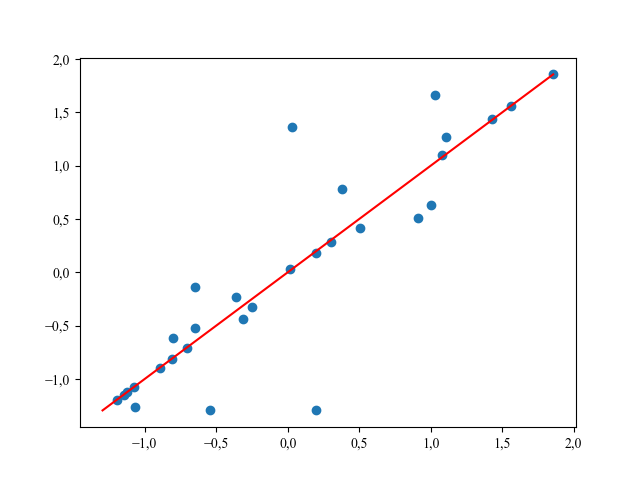
Степень полинома = 24,  
R2 значение модели = 0.8068441921670091



Степень полинома = 25,  
R2 значение модели = 0.8059211969009918

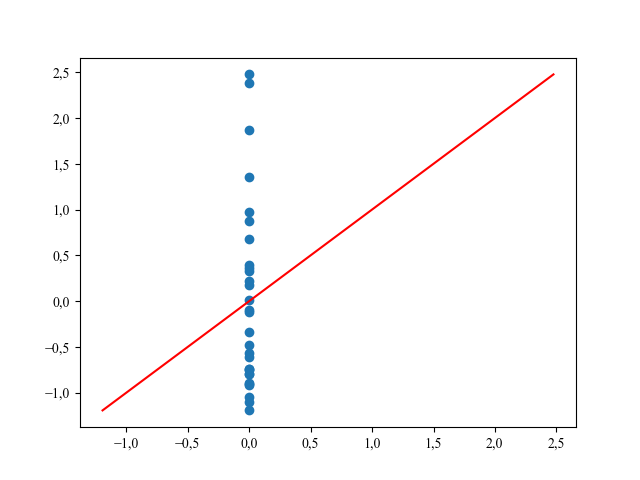


Степень полинома = 26,  
R2 значение модели = 0.80516038359366

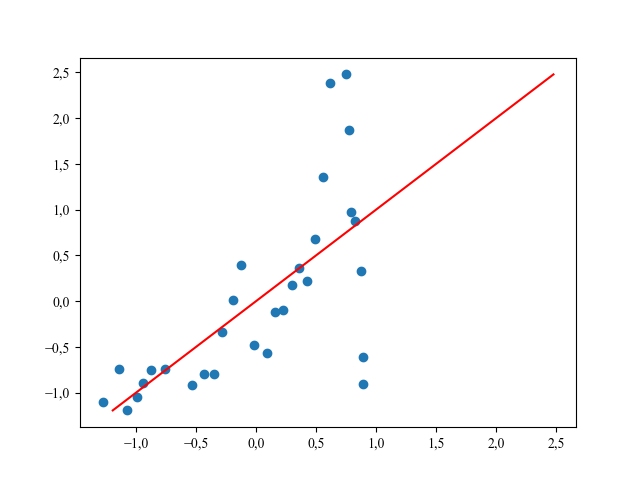


Все равно R2 счет мал.  
  
 Если смотреть на график распределения под углом в 90 градусов, то можно предположить, что это полином степени N.  
 Попробуем сменить зависимую свободную переменные местами и построить модель.

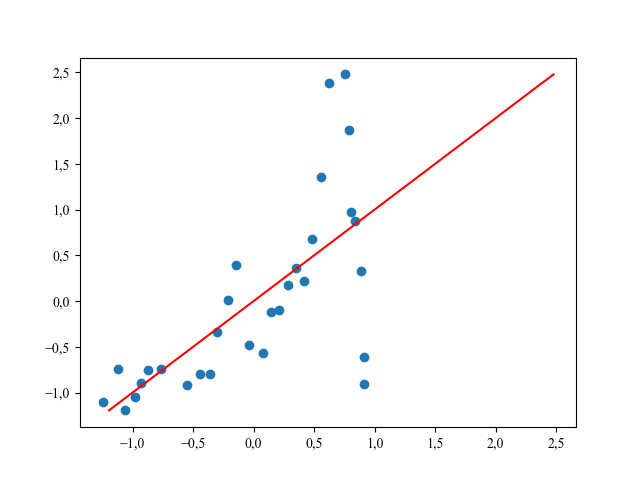
Степень полинома = 0,  
R2 значение модели = 0.0



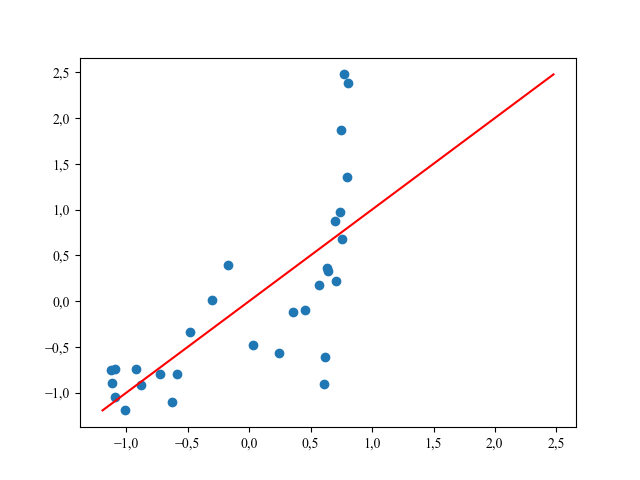
Степень полинома = 1,  
R2 значение модели = 0.4757456570052697



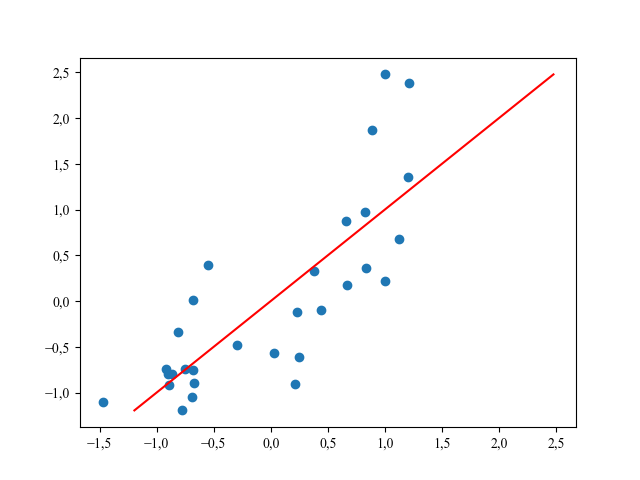
Степень полинома = 2,  
R2 значение модели = 0.4759605114246941



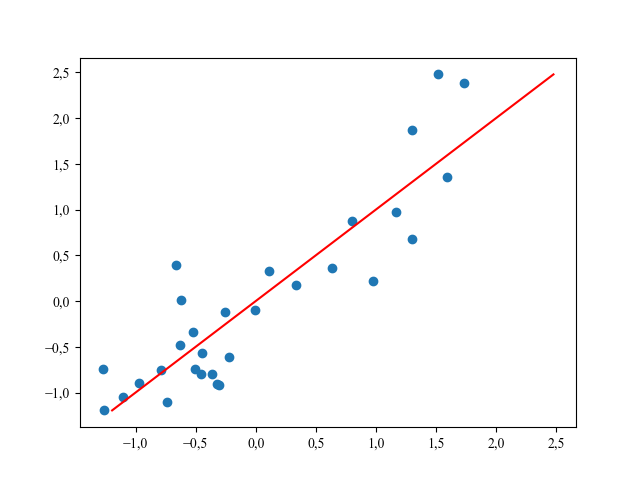
Степень полинома = 3,  
R2 значение модели = 0.5337188726422406



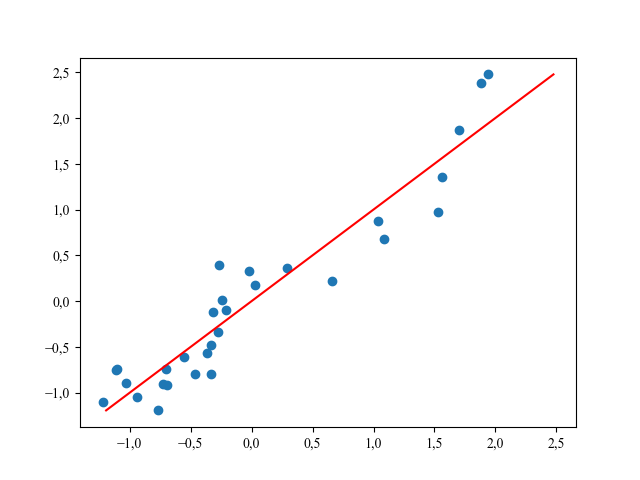
Степень полинома = 4,  
R2 значение модели = 0.6396616270252964



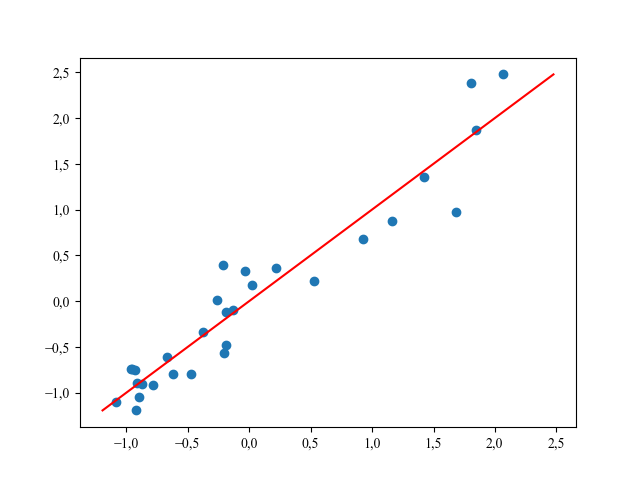
Степень полинома = 5,  
R2 значение модели = 0.7950168242978981



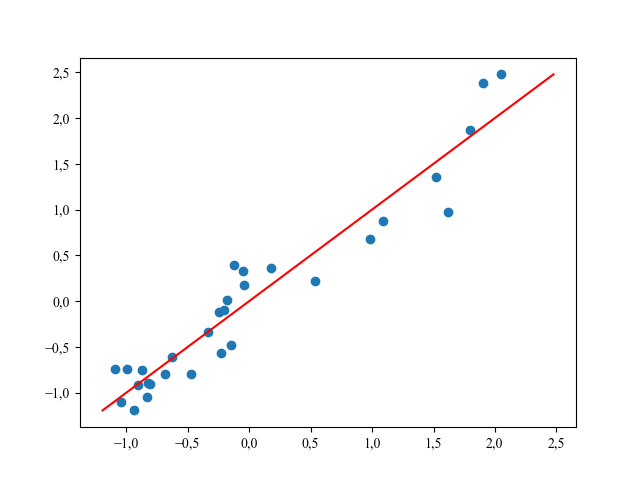
Степень полинома = 6,  
R2 значение модели = 0.9007987437501633



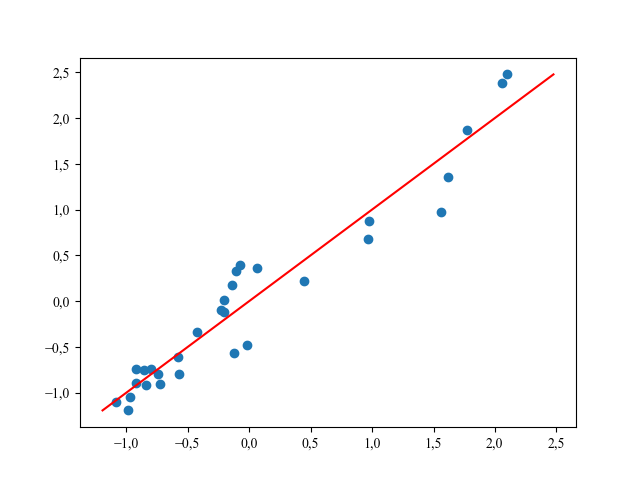
Степень полинома = 7,  
R2 значение модели = 0.9173394112012301



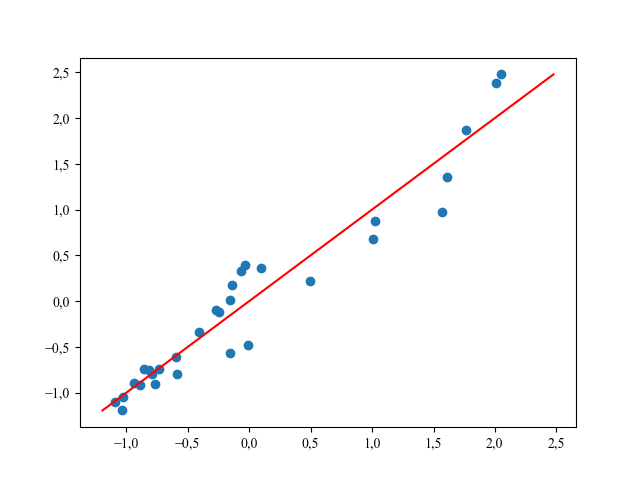
Степень полинома = 8,  
R2 значение модели = 0.9217771316770653



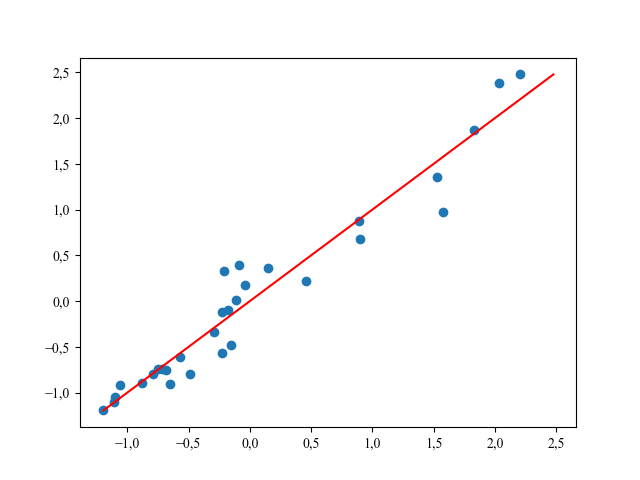
Степень полинома = 9,  
R2 значение модели = 0.9303454681250569



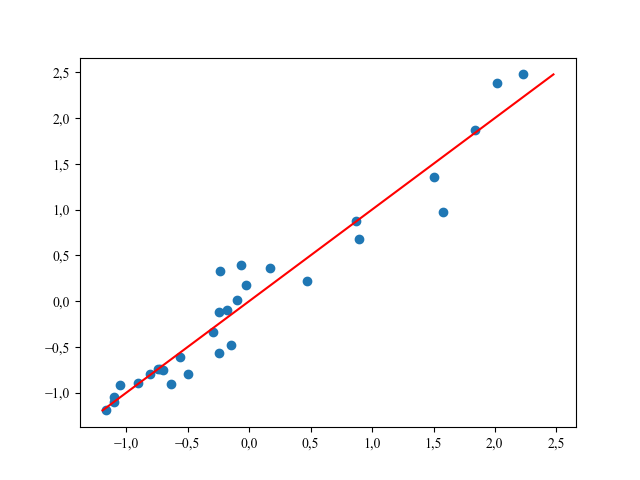
Степень полинома = 10,  
R2 значение модели = 0.9318227920993224



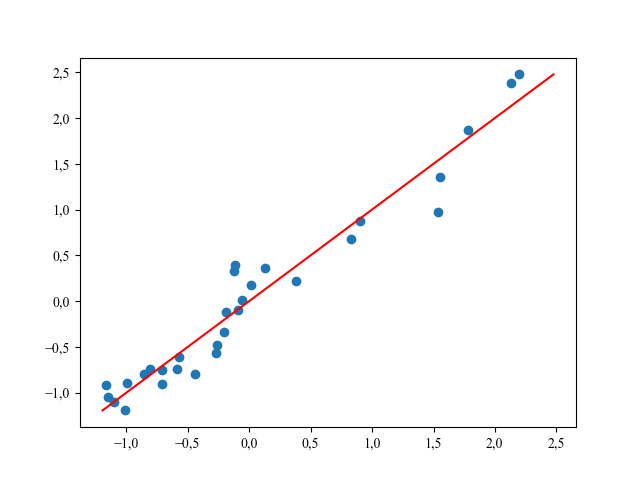
Степень полинома = 11,  
R2 значение модели = 0.9412010128138524



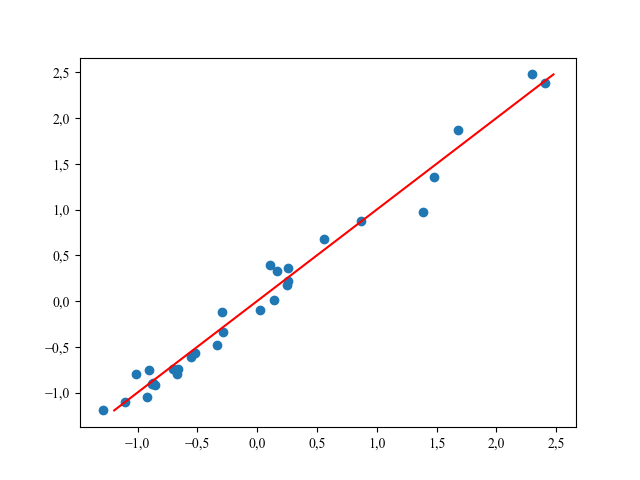
Степень полинома = 12,  
R2 значение модели = 0.9414642608655663



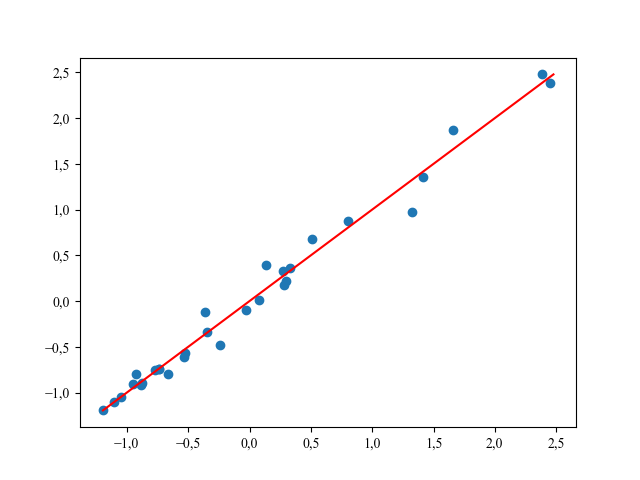
Степень полинома = 13,  
R2 значение модели = 0.9470161845737495



Степень полинома = 14,  
R2 значение модели = 0.9792328874834196

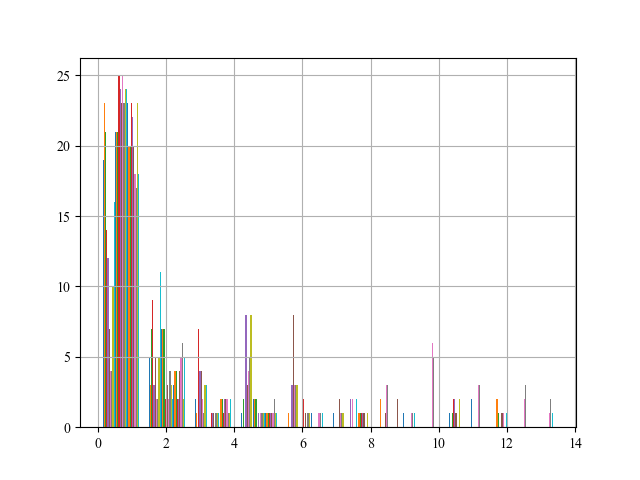


Степень полинома = 15,  
R2 значение модели = 0.983655425621807



Мы нашли модель с R2 = 0.983655425621807

Проведем тестирование модели.   
  
 Проверим остатки на нормальность визуально



K-S тест

Тест Колмогорова-Смирнова (или K-S тест) — это непараметрический статистический тест, применяемый для проверки соответствия распределения выборки заданному теоретическому распределению. Тест позволяет оценить, насколько эмпирическое распределение данных совпадает с нормальным распределением или с любым другим теоретическим распределением.  
  
Основные этапы алгоритма теста Колмогорова-Смирнова:  
Сбор данных. Получаем выборку, для которой нужно проверить соответствие распределению.  
  
Определение теоретического распределения. Выбираем теоретическое распределение, с которым будем сравнивать данные (например, нормальное, равномерное и т.д.).  
  
Построение эмпирической функции распределения (ЭФР):  
  
Вычисляем кумулятивные частоты значений в выборке, чтобы построить эмпирическую функцию распределения.  
  
Построение теоретической функции распределения (ТФР):  
  
На основе выбранного теоретического распределения рассчитываем его кумулятивную функцию распределения для каждого значения в выборке.  
  
Вычисление статистики Колмогорова-Смирнова:  
  
Определяем максимальное отклонение между эмпирической и теоретической функциями распределения: D = max | F\_эмп(x) - F\_теор(x) |, где F\_эмп(x) — значение эмпирической функции распределения, F\_теор(x) — значение теоретической функции распределения для каждого значения x в выборке.  
  
Сравнение с критическим значением:  
  
Полученное значение D сравнивается с критическим значением для заданного уровня значимости (обычно 0,05 или 0,01), которое зависит от объема выборки.  
Если D превышает критическое значение, гипотеза о совпадении распределений отклоняется.  
  
Интерпретация результатов:  
  
Если D меньше критического значения: гипотеза о том, что данные следуют теоретическому распределению, не отклоняется.  
Если D больше критического значения: гипотеза о соответствии распределению отклоняется, что говорит о значительных отклонениях данных от выбранного распределения.  
  
Тест Колмогорова-Смирнова часто используется для проверки нормальности и других распределений. Он также применим для двухвыборочного теста, когда нужно проверить, принадлежат ли две выборки одному и тому же распределению.

Jarque-Bera

Тест Джарка-Бера (Jarque-Bera) — это статистический тест, используемый для проверки нормальности распределения данных. Он основывается на оценке асимметрии (сместности) и эксцесса (пиковости) распределения, чтобы определить, насколько распределение данных отличается от нормального.  
  
Основные этапы алгоритма теста Джарка-Бера:  
Сбор данных. Получаем выборку, для которой нужно проверить нормальность.  
  
Вычисление параметров:  
  
n: объем выборки.  
Среднее значение выборки.  
Стандартное отклонение выборки.  
  
  
Рассчитываем асимметрию и эксцесс:  
  
Асимметрия (skewness). Измеряет, насколько данные симметричны относительно среднего. Формула: S = (1/n) \* сумма [(x\_i - среднее) / стандартное отклонение]^3.  
Эксцесс (kurtosis). Показывает, насколько распределение «пикообразно» или «плосковершинно». Формула: K = (1/n) \* сумма [(x\_i - среднее) / стандартное отклонение]^4 - 3.  
  
Расчет статистики теста Джарка-Бера: JB = (n/6) \* (S^2 + (K^2)/4). Чем больше значение JB, тем сильнее отклонение от нормальности.  
  
Сравнение с критическим значением:  
  
Полученное значение статистики JB сравнивается с критическим значением из распределения хи-квадрат с 2 степенями свободы на выбранном уровне значимости (обычно 0,05).  
  
Если JB превышает критическое значение, то гипотеза нормальности отклоняется.  
  
Интерпретация результатов:  
Если JB меньше критического значения: гипотеза о нормальности не отклоняется, и можно предположить, что данные распределены нормально.  
Если JB больше критического значения: гипотеза о нормальности отклоняется, что говорит о наличии значительной асимметрии или отклонений от нормальной формы распределения.  
  
Этот тест полезен для предварительного анализа данных и проверки предположения о нормальности, что важно во многих статистических методах и эконометрических моделях.

Статистика Jarque-Bera: 1.4622846121728859

p-значение: 0.4813588168106787

Данные распределены нормально

Shapiro-Wilk

Статистика Shapiro-Wilk: 0.9494766810692478

p-значение: 0.1636875606933047

Распределение данных похоже на нормальное

Helwig

Шаг 1: Сортируем данные и определяем размер выборки

Шаг 2: Оценка среднего и стандартного отклонения

Шаг 3: Вычисляем эмпирическую функцию распределения (ЭФР)

Шаг 4: Строим теоретическую нормальную функцию распределения (НФР)

Шаг 5: Вычисляем максимальное отклонение между ЭФР и НФР

Вывод результата

Максимальное отклонение (D): 0.11388137902302198

Нет оснований отвергнуть гипотезу о нормальности на уровне значимости 0.05.

Сравнение тестов

Сравнение методов согласия Хельвига, Шапиро-Вилька и Джарка-Бера (Jarque-Bera) полезно для выбора подходящего теста для проверки нормальности распределения данных. Каждый из этих методов имеет свою область применения и особенности, которые могут быть полезны в разных контекстах.  
  
1. Тест Хельвига  
Цель: Метод Хельвига основан на анализе корреляций и используется для оценки согласия признаков, особенно в социально-экономических и психометрических исследованиях.  
Применение: Обычно применяется для оценки многомерного согласия признаков или при проведении факторного анализа.  
Преимущества:  
Хорошо подходит для многомерных данных, поскольку анализирует согласие между несколькими переменными.  
Позволяет оценить общую структуру корреляций между признаками, что важно для анализа взаимозависимости.  
Недостатки:  
Не подходит для проверки нормальности распределения данных.  
Может требовать больших выборок для корректного анализа многомерных данных.  
  
2. Тест Шапиро-Вилька  
Цель: Проверка нормальности распределения данных в выборке.  
Применение: Часто используется для малых и средних выборок (до 2000 наблюдений), чтобы оценить, насколько распределение данных близко к нормальному.  
Преимущества:  
Очень чувствителен к отклонениям от нормальности, особенно в малых выборках.  
Является одним из самых мощных тестов для проверки нормальности, так как учитывает порядок значений в выборке.  
Недостатки:  
Может давать ложные результаты для больших выборок (более 2000 наблюдений), так как становится излишне чувствительным к малейшим отклонениям.  
Не подходит для многомерных данных, так как используется для одномерного распределения.  
  
3. Тест Джарка-Бера (Jarque-Bera)  
Цель: Проверка нормальности распределения путем оценки асимметрии (skewness) и эксцесса (kurtosis).  
Применение: Часто применяется для данных больших объемов, особенно в эконометрических и финансовых исследованиях.  
Преимущества:  
Хорошо подходит для больших выборок, так как рассчитывается на основе асимметрии и эксцесса, которые более устойчивы в больших объемах данных.  
Удобен для случаев, когда нужны простые показатели нормальности (асимметрия и эксцесс).  
Недостатки:  
Менее чувствителен для малых выборок, так как асимметрия и эксцесс могут быть нестабильными.  
Не учитывает порядок значений в выборке, что делает его менее точным для малых выборок.  
  
Вывод:  
Для малых выборок (до 2000 наблюдений) тест Шапиро-Вилька наиболее подходит для проверки нормальности, поскольку он высокочувствителен к отклонениям и учитывает порядок значений.  
Для больших выборок (более 2000 наблюдений) тест Джарка-Бера предпочтителен, так как он основан на асимметрии и эксцессе, что стабильно в больших объемах данных.  
Тест Хельвига лучше использовать, когда требуется оценить согласие нескольких переменных, а не нормальность, так как он лучше подходит для анализа многомерных зависимостей.  
Таким образом, выбор метода зависит от цели исследования, объема выборки и характеристик данных.