**Задание 3.** Методы машинного обучения с учителем: деревья решений, ансамбли деревьев решений

1) Решение задачи классификации

– Загрузите набор данных по раку молочной железы с помощью функции load\_breast\_cancer из scikit-learn. Разбейте его на обучающий и тестовый наборы.

– Создайте экземпляр модели DecisionTreeClassifier и обучите его на обучающем наборе данных. Оцените правильность полученной модели на обучающем и тестовом наборах.

– Повторите эксперимент, используя значения параметра max\_depth равные 1, 2, … 10. Постройте график зависимости правильности модели на обучающем и тестовом наборах данных от значения max\_depth.

– Обучите модель RandomForestClassifier, используя значения параметра n\_estimators равные 5, 10, … 50. Постройте график зависимости правильности модели на обучающем и тестовом наборах данных от значения n\_estimators.

2) Решение задачи регрессии

– Сгенерируйте синтетический набор данных:

np.random.seed(0)

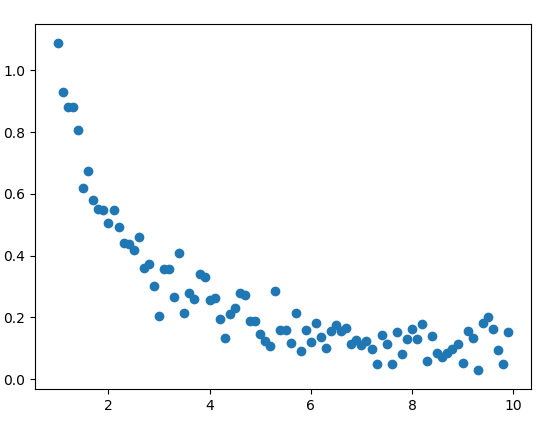
X = np.arange(1.0, 10.0, 0.1).reshape(-1, 1)

Y = np.array([max(0, (1 / x) + np.random.normal(scale = 0.05)) for x in X])

– Постройте точечную диаграмму по полученным данным:

plt.scatter(X, Y)

plt.show()



– Разбейте сгенерированный набор данных на обучающий и тестовый наборы.

– Создайте экземпляр модели LinearRegression и обучите его на обучающем наборе данных. Оцените качество полученной модели на обучающем и тестовом наборах используя метрику или среднюю квадратическую ошибку.

– Создайте экземпляр модели DecisionTreeRegressor и обучите его на обучающем наборе данных. Оцените качество полученной модели на обучающем и тестовом наборах используя метрику или среднюю квадратическую ошибку и сравните результат с полученным на предыдущем шаге.

– Загрузите набор данных Boston Housing с помощью функции load\_boston из scikit-learn. Разбейте его на обучающий и тестовый наборы. Обучите модель DecisionTreeRegressor, используя значения параметра max\_depth равные 1, 2, … 10. Постройте график зависимости или средней квадратической ошибки на обучающем и тестовом наборах данных от значения max\_depth.

– Обучите модель RandomForestRegressor, используя значения параметра n\_estimators равные 5, 10, … 50. Постройте график зависимости или средней квадратической ошибки на обучающем и тестовом наборах данных от значения n\_estimators.