**Задание 2.** Методы машинного обучения с учителем: линейные модели, ядерный метод опорных векторов

1) Решение задачи регрессии

– Загрузите набор данных Boston Housing с помощью функции load\_boston из scikit-learn. Разбейте его 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.

– Обучите модель Lasso, используя значения параметра alpha равные 0,01, 0,1, 1, 10 и 100. Постройте график или таблицу зависимости или средней квадратической ошибки на обучающем и тестовом наборах данных от значения alpha. Сравните (в виде графика или таблицы) значения коэффициентов обученной модели для значений alpha равных 0,1 и 10.

– Обучите модель Ridge, используя значения параметра alpha равные 0,01, 0,1, 1, 10 и 100. Постройте график или таблицу зависимости или средней квадратической ошибки на обучающем и тестовом наборах данных от значения alpha. Сравните (в виде графика или таблицы) значения коэффициентов обученной модели для значений alpha равных 0,1 и 10.

– Сравните (в виде графика или таблицы) значения коэффициентов обученных моделей Lasso и Ridge для alpha=10.

2) Решение задачи классификации

– Сгенерируйте синтетический набор данных:

np.random.seed(0)

X = np.random.rand(50, 2)

X = np.append(X, np.random.rand(50, 2) + np.array((3.0, 3.0)), axis = 0)

X = np.append(X, np.random.rand(50, 2) + np.array((0.0, 2.0)), axis = 0)

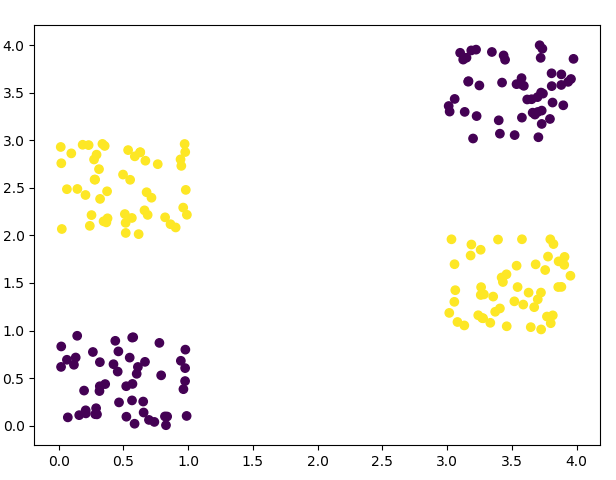
X = np.append(X, np.random.rand(50, 2) + np.array((3.0, 1.0)), axis = 0)

Y = np.append(np.zeros(100), np.ones(100), axis = 0)

– Постройте точечную диаграмму по полученным данным:

pyplot.scatter(X[:,0], X[:,1], c=Y)

pyplot.show()



– Разбейте сгенерированный набор данных на обучающий и тестовый наборы.

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

– Создайте экземпляр модели SVC и обучите его на обучающем наборе данных. Оцените правильность полученной модели на обучающем и тестовом наборах и сравните результат с полученным на предыдущем шаге.

– Загрузите набор данных по раку молочной железы с помощью функции load\_breast\_cancer из scikit-learn. Разбейте его на обучающий и тестовый наборы. Обучите модель SVC, используя значения параметра gamma равные 0,0001, 0,001, 0,01, 0,1 и 1. Постройте график или таблицу зависимости правильности модели на обучающем и тестовом наборах данных от значения gamma.

– Обучите модель SVC, используя значения параметра C равные 0,001, 0,01, 0,1, 1 и 10. Постройте график или таблицу зависимости правильности модели на обучающем и тестовом наборах данных от значения C.