**Задание 6.** Оценка и улучшение качества модели

1) Перекрестная проверка

– Загрузите набор данных по раку молочной железы с помощью функции load\_breast\_cancer из scikit-learn.

– Создайте экземпляр модели LogisticRegression и рассчитайте для него среднюю правильность перекрестной проверки на наборе данных, указав значение параметра cv равным 5.

2) Решетчатый поиск

– 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 SVC и задайте сетку параметров, используя значения gamma равные 0,0001, 0,001, 0,01, 0,1, 1 и значения C равные 0,001, 0,01, 0,1, 1, 10.

– Воспользуйтесь GridSearchCV для поиска лучших параметров на обучающем наборе данных, указав значение параметра cv равным 5. Выведите наилучшие значения параметров и кросс-валидационной правильности. Оцените правильность наилучшей модели на тестовом наборе данных.

3) Метрики для бинарной классификации

– Загрузите набор данных по раку молочной железы с помощью функции load\_breast\_cancer из scikit-learn. Обозначьте примеры с злокачественными опухолями как положительные, а примеры с доброкачественными опухолями – как отрицательные:

cancer = load\_breast\_cancer()

X = cancer.data

Y = cancer.target == 0

– Отбросьте часть положительных примеров таким образом, чтобы их количество составляло около 10% от общего количества примеров:

X\_positive = X[Y]

Y\_positive = Y[Y]

X\_negative = X[Y == False]

Y\_negative = Y[Y == False]

n\_positive = int(len(X\_negative) / 9)

X\_positive = X\_positive[0:n\_positive]

Y\_positive = Y\_positive[0:n\_positive]

X = np.append(X\_negative, X\_positive, axis = 0)

Y = np.append(Y\_negative, Y\_positive, axis = 0)

– Разбейте полученный набор данных на обучающий и тестовый наборы.

– Создайте экземпляр модели DummyClassifier и обучите его на обучающем наборе данных, указав значение параметра strategy равным “most\_frequent”. Оцените правильность и полноту полученной модели на обучающем и тестовом наборах.

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

– Для обученной модели постройте ROC-кривую и рассчитайте значение AUC на тестовом наборе данных.