

## 1. Почему более эффективно использовать F-меру, а не точность?

*Precision(точность)* - это доля объектов, названных классификатором положительными и при этом действительно являющимися положительными. Также в задачах классификации используется такая мера, как *recall(полнота)*. Это доля тех объектов положительного класса, которые действительно принадлежат этому классу и сеть правильно их отнесла.

Хорошей будет та модель, у которой максимальная точность и полнота одновременно, но в реальной жизни очень сложно этого достигнуть. Для этого и была создана F-мера, которая является гармоническим средним между точностью и полнотой. Таким образом, она объединяет в себе информацию о точности и полноте и позволяет лучше оценить модель.

$$F = 2 \frac{Precision \times Recall}{Precision + Recall}$$

## 2. Какие виды слоев используются в АРТ-сетях? Какой функционал они выполняют?

*Слой сравнения* - число нейронов на нем фиксировано и равно размерности входного вектора. На данном слое осуществляется сравнение значения слоя распознавания и текущего входного образца. Для этого входной вектор преобразуется в вектор *S*, который передается в весовую матрицу действительных чисел. На выходе каждого нейрона слоя распознавания формируется ожидаемый вектор *V* (типичный представитель данного класса) и из них выбирается образец с наибольшим соответствием. Затем *S* и *V* сравниваются с так называемым параметром соответствия. Если схожести не обнаружено, то на слое распознавания создается новый нейрон.

*Слой распознавания* (выходной слой/соревновательный слой) - число нейронов переменное, каждому нейрону соответствует некоторый класс. Первоначально на слое есть только один нейрон. Данный слой сопоставляет каждому входному вектору класс. Если подходящего класса нет, образуется новый класс (к слою добавляется новый нейрон).

### **3. Какие способы представления текста Вы знаете? Какой используете в работе?**

Для обработки любой текстовой информации текст делится на токены, а затем производится векторизация одним из следующих способов:

- числовое кодирование (ставит в соответствие токену число. Это может быть частота встречаемости в тексте, UTF-8 код и т.д.)
- One hot encoding (ставит в соответствие каждому токену вектор длины количества токенов, все элементы которого равны 0, кроме того, который соответствует токену)
- Плотные векторные представления (ставит каждому токену в соответствие вектор, размерность которого меньше количества токенов и кроме 0 и 1 присутствуют другие числа.

В лабораторной работе используется числовое кодирование.

### **4. Для чего может быть использовать noise\_shape в слое Dropout?**

Параметр noise\_shape представляет собой форму маски разряжения. Данный параметр позволяет использовать на всех шагах одинаковую маску, если указать noise\_shape = (batch\_size, 1, features) при форме входа (batch\_size, timesteps, features).