



Министерство образования и науки Российской  
Федерации

Федеральное государственное автономное  
образовательное учреждение высшего образования  
«Уральский федеральный университет имени  
первого Президента России Б. Н. Ельцина»

Основы анализа данных с Python и SQL

Отчет

По лабораторной работе №5

**Классификация изображений сверточными моделями**

Студент: Манакин Андрей

г. Екатеринбург

2024 г.

Ответы на вопросы из ноутбука:

**Сделайте вывод о сбалансированности датасета:**

Датасет можно назвать сбалансированным, потому что в него входит 10 классов, по 6 тыс. изображений в каждом и они распределены равномерно, исходя из гистограммы выше

**Исходя из этого, какие метрики будем применять для оценки качества модели**

Т.к датасет имеет сбалансированное распределение, для оценки качества модели можно использовать такие стандартные метрики как: Accuracy и Cross-Entropy Loss

**Сделайте разделение на train и val, зачем так делать?**

Разделение на Train и Val необходимо для того, чтобы оценить как модель работает с незнакомыми ей данными, на трейне модель обучается, затем проверяется на валидейта, если аккуратность для трейна растёт, а для валидейта стоит на месте, значит модель переобучается

**Создайте Dataloader'ы и настройте гипер параметры, оптимизатор и функцию потерь, какой оптимизатор будете использовать и почему, а какую loss-функцию?**

В качестве оптимизатора я выберу Adam, так как он эффективно справляется с регуляризацией. loss-функцию буду использовать CrossEntropyLoss(), так как она хорошо подходит для задач многоклассовой классификации

**Сделайте выводы о качестве обучения**

Примерно с 20-ой эпохи точность валидации перестала расти, а местами начала уменьшаться, в то время, как точность на тренируемой выборке продолжает расти, это говорит о том, что модель переобучается и с этим что-то нужно делать.

**Новая модель:**

Для своей новой модели я добавил слой аугментации, увеличил количество фильтров в сверточных слоях для извлечения большего кол-во признаков из изображений, добавил дропауты, для предотвращения переобучения, а также слои

нормализации батча, стабилизирующие и ускоряющие процесс обучения, так же я увеличил количество эпох до 150, что позволило моей модели получить необходимую величину Ассигасы для тестового набора данных