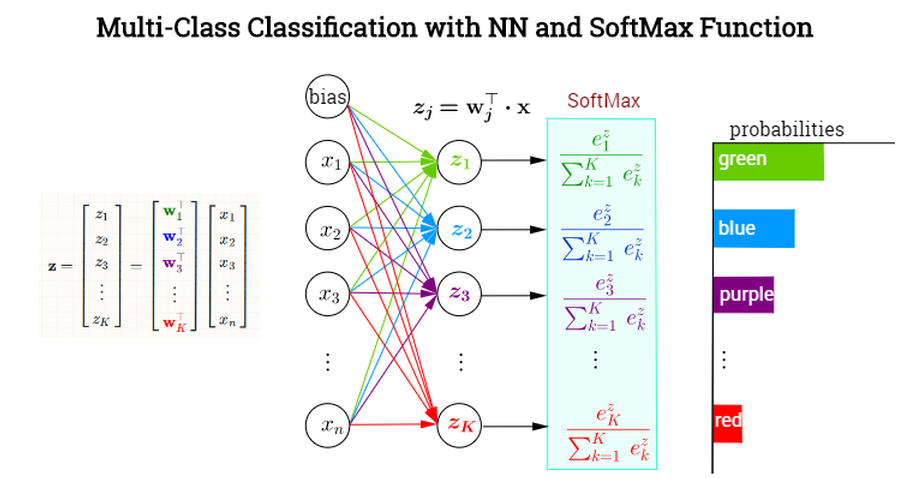
Нейросеть состоит из нескольких слоев нейронов, и будет тренироваться через обратное распространение ошибки.

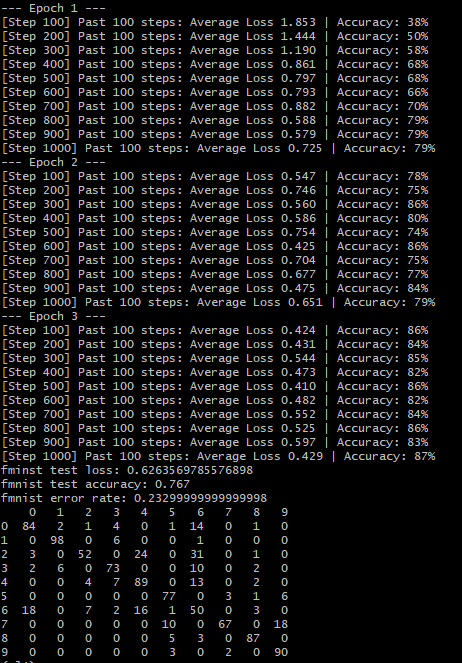
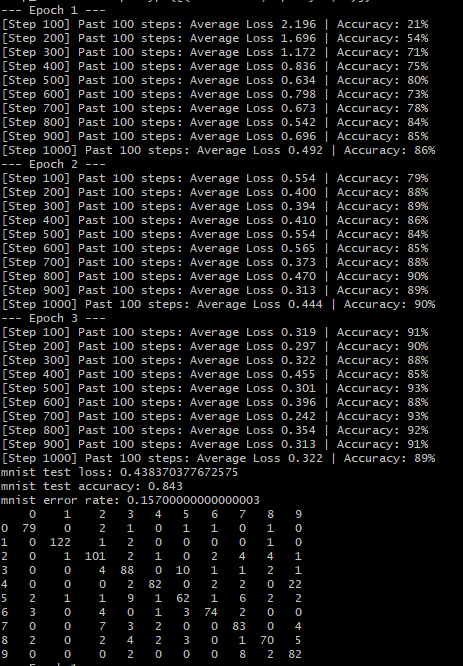
Adam — adaptive moment estimation, ещё один оптимизационный алгоритм (поиск минимума функции). Он сочетает в себе и идею накопления движения («Если мы некоторое время движемся в определённом направлении, то, вероятно, нам следует туда двигаться некоторое время и в будущем», накапливаются значения градиента) и идею более слабого обновления весов для типичных признаков (проверяется, как часто меняется градиент). Нетипичные признаки (встречающиеся реже) обновляют веса сильнее. Используется для оптимизации весов.

Convolution layer - сверточный слой. Берется небольшой фильтр – матрица чисел (в нашем случае 3 на 3) и применяется к каждой группе пикселей исходного изображения (т е движется по изображению). В зависимости от того, с какой частотой проходит фильтр, на выходе может получиться матрица (изображение) разного размера – от исходного, если добавить паддинги из нулей по краям, до совсем небольшого. Из полученного изображения можно выделить только нужные части (к примеру, бинаризовать в зависимости от порогового значения пикселей), возможно применение нескольких фильтров таким образом. Так на изображении выделяются некоторые фрагменты. Потом могут сопоставляться именно фрагменты, а не все сразу.

Max pooling layer – слой, который работает аналогично, но берет максимальное значение из обрасти (2 на 2 или 3 на 3 обычно). Борется с переобучением, также позволяет сократить вычисления (сохраняется паттерн при уменьшении объема данных). Также возможно использовать average pooling – среднее значение по области.

SoftArgMax (also known as softargmax or normalized exponential function) – функция, которая по ненормализованным данным может посчитать вероятность каждого ответа. В нашем случае предварительно превращается массив изображений (для этого также есть оптимизируемые веса).





Первый скриншот на mnist, второй – fmnist (ожидаемо результаты ухудшились).

Подобран результат в 8 фильтров, матрица 3х3, пулинг осуществляется 4 клетки (2х2) в 1.

В матрицах неточности можно явно заметить, что числа в диагонали явно превышают прочие числа (классификация производится хорошо).

Матрица классификации изображений соответственно – диагональ единичек (ожидаемо).