MNIST

**Baseline 1.0**

784 -> 625 dense sigmoid

625 -> 10 dense softmax

SGD default

categorical\_crossentropy

initialization: np.random.randn \* 0.01

Test accuracy: 0.92

No sign of overfitting

Slow: converge after 80 epochs

Based on Baseline 1.0

784 -> 625 dense sigmoid => relu

Test performance is much better. However, limited by the training speed of SGD

**Baseline 1.1:**

Based on Baseline 1.0

784 -> 625 dense sigmoid => relu

SGD => rmsprop default

Test performance 0.981 or 0.982

Converge around 10th epochs.

Train accuracy 0.999 (sign of overfitting)

Based on Baseline 1.1

784 -> 625 dense sigmoid => + dropout (0.3)

Test performance 0.984

Converge around 10th epochs.

Train accuracy 0.995 (sign of overfitting)

Based on Baseline 1.1

784 -> 625 dense sigmoid => + dropout (0.7)

Test performance 0.98

Converge around 10th epochs.

Train accuracy 0.978, even worse than test performance

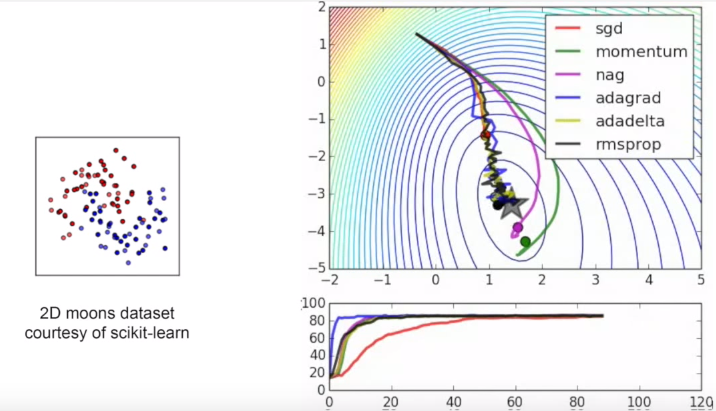
Based on Baseline 1.1

784 -> 625 dense sigmoid => + dropout (0.5)

Test performance 0.983

Converge around 10th epochs.

Train accuracy 0.990 (sign of overfitting)



Adagrad seems to converge very fast

Takeaway:

If your data has structure, your model should reflect it.