# Regularization Tests

## Weight Constraints:

Lambda is the limit of the weights, $w \in [-\lambda, \lambda]$.

All tests are shown in the last epoch (30).

| SGD optimizer with eta 0.5, cross-entropy cost function, regularization Weights Constraints | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Lambda | 0.25 | 0.5 | 0.75 | 1.0 | 1.25 | 1.5 | 1.75 | 2.0 | 2.25 |
| Negative Log Loss | 0.1374 | 0.0622 | 0.0424 | 0.0344 | 0.0309 | 0.0278 | 0.0271 | 0.0267 | 0.0237 |
| Training Accuracy | 87.85% | 91.91% | 93.87% | 95.19% | 95.86% | 96.17% | 96.76% | 96.83% | 97.05% |
| Validation Accuracy | 90.7% | 93.58% | 94.98% | 95.62% | 96.08% | 96.4% | 96.36% | 96.54% | 96.78% |

## L1

Lambda is L1 regularization factor, $w^{t+1} = w^t - \frac{\eta \cdot \lambda}{n} \cdot sign(w^t) - \frac{\eta}{m} \cdot \nabla w^t$, for learning rate $\eta$, $n$ is the number of training examples and $m$ is the size of the batch.

All tests are shown in the last epoch (30).

| SGD optimizer with eta 0.5, cross-entropy cost function, regularization L1 | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Lambda | 0 | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 | 0.6 | 0.7 | 0.8 | 0.9 |
| Negative Log Loss | 0.0229 | 0.0229 | 0.0222 | 0.0238 | 0.0228 | 0.0229 | 0.0232 | 0.0235 | 0.0228 | 0.023 |
| Training Accuracy | 97.59% | 97.89% | 97.75% | 97.67% | 97.7% | 97.71% | 97.56% | 97.59% | 97.66% | 97.46% |
| Validation Accuracy | 96.74% | 97.2% | 96.94% | 96.82% | 96.92% | 96.88% | 96.9% | 96.82% | 96.84% | 97.06% |

## L2

Lambda is L2 regularization factor, $w^{t+1} = \left(1 - \frac{\eta \cdot \lambda}{n}\right) \cdot w^t - \frac{\eta}{m} \cdot \nabla w^t$, for learning rate $\eta$, $n$ is the number of training examples and $m$ is the size of the batch.

All tests are shown in the last epoch (30).

| SGD optimizer with eta 0.5, cross-entropy cost function, regularization L2 | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Lambda | 0 | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 | 0.6 | 0.7 | 0.8 | 0.9 |
| Negative Log Loss | 0.0244 | 0.0236 | 0.0229 | 0.0238 | 0.0228 | 0.025 | 0.0223 | 0.0238 | 0.0236 | 0.0225 |
| Training Accuracy | 97.56% | 97.67% | 97.8% | 97.7% | 97.75% | 97.62% | 97.5% | 97.73% | 97.7% | 97.52% |
| Validation Accuracy | 96.82% | 96.78% | 97.04% | 96.74% | 97.06% | 96.6% | 97.12% | 96.68% | 96.86% | 97.2% |

**No-improvement-in-n regularization:**

Dividing $\eta$ by 2, when there is no improvement for $n$ epochs. Minimum $\eta$ is $\frac{\eta_0}{128}$.

Also, the algorithm will stop if there is no improvement in 5 epochs.

| **SGD optimizer with eta 0.5, cross-entropy cost function, regularization no-improvement-in-n** | | | | | | |
|---|---|---|---|---|---|---|
| n | 0 | 1 | 2 | 3 | 4 | 5 |
| Epoch | 30 | 30 | 30 | 30 | 30 | 30 |
| eta | $\eta_0$ | $\frac{\eta_0}{16}$ | $\frac{\eta_0}{4}$ | $\eta_0$ | $\eta_0$ | $\eta_0$ |
| Negative Log Loss | 0.024 | 0.0236 | 0.0234 | 0.237 | 0.0233 | 0.0232 |
| Training Accuracy | 97.58% | 97.52% | 97.68% | 97.77% | 97.7% | 97.71% |
| Validation Accuracy | 96.72% | 97.02% | 96.8% | 96.82% | 97.02% | 97.06% |