Linear Classification.

Setting Delta. Note that we brushed over the hyperparameter ΔΔ and its setting. What value should it be set to, and do we have to cross-validate it? It turns out that this hyperparameter can safely be set to Δ=1.0 in all cases. The hyperparameters ΔΔ and λλ seem like two different hyperparameters, but in fact they both control the same tradeoff: The tradeoff between the data loss and the regularization loss in the objective. The key to understanding this is that the magnitude of the weights WW has direct effect on the scores (and hence also their differences): As we shrink all values inside WW the score differences will become lower, and as we scale up the weights the score differences will all become higher. Therefore, the exact value of the margin between the scores (e.g. Δ=1, or Δ=100) is in some sense meaningless because the weights can shrink or stretch the differences arbitrarily. Hence, the only real tradeoff is how large we allow the weights to grow (through the regularization strength λ).