

Marking Guidelines for LFD 2nd assignment 2006

Kian Ming Adam Chai

December 27, 2006

Marking Scheme

70 - 100	A	Excellent
60 - 69	B	Very Good
50 - 59	C	Good
40 - 49	D	Satisfactory
35 - 39	E	Marginal Fail
25 - 34	F	Clear Fail
00 - 24	G	Bad Fail

1st Question [25]

Description of Individual Attributes [8]

- range [2]
- mean / median [2]
- std/var [2]
- histogram [2]
- Clusters/quantization [2]
- Description: Gaussian, skewed etc. [2]
- Provide numbers [2]

Target versus attributes dependency [8]

- negative / decreasing [2]
- Correlation [2]
- linear / non-linear / quadratic [2]
- Note presense/absence of noise [1]
- Provide numbers [2]
- Others, e.g.
 - Hyperplane [1]
 - Variance of function [1]
 - Local versus global variation [1]
 - Distribution of targets only [1]

Correlation between Attributes [4]

- Correlation [2]
- PCA /eigen-decomposition [2]
- Implication of results [2]
- Provide numbers [1]
- Note presense/absence of noise [1]

Comparison between different data sets [5]

- Notes scale / range differences or cluster around different regions [2]
- Presence/absence of outliers [1]
- Implications of results [2]
- What is the ideal case? [1]
- Others [1 ea]

Bonus • Why do we need to visualise [2]

2nd Question [20]

Reasons for rescaling and code [10]

- Rescaling based on perceived distribution of variables [2]
- Data have different scale/range [2]
- Prevent any features from dominating [2]
- Good range for tanh/transfer function [2]
- Gradient in good range [2]
- Prevent Saturation [2]
- Range/distribution of initial weights [2]
- Investigates what `netlab` assumes [2]
- Correct code: (mean/std) or (min,max) [4]
- Add a 1 for zero std: No credit given, since I thought should add something like 10^{-3} . [0]

Cannot use pseudocode. What is to be done [10]

- Rerun code would give different means / variance from training data [2]
- Require same transformation for consistency [3]
- Need to use the same mean /variance as training data [2]
- Actual testing may not have entire set for statistics [1]
- Training should use training set only anyway [1]
- Others [1 or 2 ea]

3rd Question [35]

Choice of structure [10]

- Number of outputs [1]
- Number of inputs [1] and why? [1]
- Number of hidden units [1] why? [1]
- Nature of hidden unit transfer function [2]
- Linear output units [1] and why [1]
- Sum-squared error /Gaussian error / absolute error [1] and why [1]
- Initialisation of weights [2]
- Choice of optimisation method, e.g. scg [2]; if gradient descent [only 1]

Process: No averaging of results [20]

- ≥ 8 no of hidden units [2]
- 1 or 2 hidden units [2]
- Some others between the two range [2]
- Some variation of the following process, with justification if very different
 - For different no of hidden units, 2 different runs, with largish iterations [2]
 - Additional random seeds for 1 or 2 best sizes [2]
 - Varying iteration with some justification [2] or randomly [only 1]
- Reports results on validation set [2]
- State final choice of network [1]
- Show knowledge that model selection is based on validation error [1]
- Early stopping [2]
- Other regularisation [2]
- Notes local minimums [1] and how to reduce this problem [1]
- Comment on usefulness of varying random seed [1]
- Comment on usefulness of varying training set size [1]
- Comment on choice of intervals between hidden units size [1]
- Comment on final choice of network, i.e. linearly separable etc [1]
- Comment on process approach i.e. brute force, greedy [1]
- Vary number of iteration based on validation error during training [+2]
- Uses clearly much more than 10 runs [-2]

Comments on train,val,test [5]

- Report results on test set using final network [2]
- Noting validation error $>$ train error because train on training set [1]
- Noting test error $>$ train error because train on training set [1]
- Noting validation error $>$ test error (or the other way, if their results say so) [1]
- Comment on overfitting varying with iterations [2]
- Comment on overfitting varying with no of hidden units [2]
- Comment on cross-validation or other schemes [1]
- Others: Correlation [1 ea]

4th Question [20]

Comments on train versus test [12]

- Ideally train $\equiv$ test [5]
- But if not, then validation $\equiv$ test [5]
- Others [1ea]

Comments on whether val $\equiv$ test in light of performance [8]

- Observation of whether val $\equiv$ test [2]
- Reference to results [3]
- Reference to visualisation [3]
- Others [1ea]

Bonus • Cross-validation [+2]

- Re-sampling [+2]
- Other ways of fixing [+1 or +2]