Homework ranking

Decision trees

Rank	Name	Score	Stdev	Impl	Model details
1*	Margo Kopli	93.8	-	weka	random forest, 10 trees
1	Hendrik Maarand	93.4	4.11	scikit	entropy cost, tree depth 7
2	Olga Dalton	91.9	2.05	self	features split into intervals of range 5, entropy cost
3	Margus Ernits	91.3	1.44	scikit	
4	Margo Kopli	92.7	2.01	weka	consider 9 random features
5	Ottokar Tilk	79.6	2.56	self	features split at median, misclassification cost, depth at least 9
6	Andrey Sergeev	20.2	0.26	self	

* All other implementations could benefit from random forest as well.

K Nearest Neighbours

Rank	Name	Score	Stdev	Impl	Model details
1	Hendrik Maarand	98.3	-	scikit	manhattan distance, K = 9, standardized, stratified cross-validation
2	Ago Luberg	97.2	0.60	self	euclidean distance, K = 21, standardized
3	Olga Dalton	96.7	0.76	self	euclidean distance, K = 21, standardized
4	Ottokar Tilk	92.6	0.00	self	euclidean distance, K = 1, standardized

K Nearest Neighbours

Rank	Name	Score	Stdev	Impl	Model details
1	Hendrik Maarand	98.3	-	scikit	manhattan distance, K = 9, standardized, stratified cross-validation
2	Ago Luberg	97.2	0.60	self	euclidean distance, K = 21, standardized
3	Olga Dalton	96.7	0.76	self	euclidean distance, K = 21, standardized
4	Ottokar Tilk	92.6	0.00	self	euclidean distance, K = 1, standardized

Neural Networks

Rank	Name	Score	Stdev	Impl	Model details
-	state-of-the-art	99.79	-	-	convolutional neural network [pdf]
-	best KNN	99.48	-	-	using some kind of distortion model [pdf]
1	Hendrik Maarand	97,17	-	self	learning rate: 0.01; learning rate decay: 0.99; 1 hidden layer with 349 units; 30% of data was used for validation/testing; Maximum number of epochs: 50 or 100?; Early stopping if there's been no improvement in 5 epochs. Scaled feature values (divided by max). Sigmoid hidden, Softmax output activation. Cross entropy error. Shuffle samples before each epoch
-	best linear classifier	92.4	-	-	some form of all-versus-all [pdf]
2	Olga Dalton	91.81	-	PyBrain	learning rate: 0.001; weight decay (L2): 0.01; 1 hidden layer with 75 units; 35% of data was used for validation; Maximum number of epochs: 20; Each time validation error hits a minimum, try for 3 epochs to find a better one (continueEpochs=3).

Linear Regression

Score is the average loss per point:

$$\text{score} = \frac{1}{m}\sum_{i=1}^{m}\left(h_{\boldsymbol{\theta}}(\mathbf{x}_i) - y_i\right)^2$$

Rank	Name	Score	Stdev	Impl	Model details
1	Ottokar Tilk	163.3	1.83	self	polynomial features (degree=4), normal equations with regularization, C = 100
2	Hendrik Maarand	168.5	0.22	self and scikit for cross-val	polynomial features (degree=2), normal equations, regularization didn't seem to matter
3	Olga Dalton	176.7	0.30	self	standardized features, gradient descent with learning rate 0.01
4	Margo Kopli	179.1	0.13	self	features 3, 5 and 6

Logistic Regression

Rank	Name	Score	Stdev	Impl	Model details
1	Ottokar Tilk	85.5	0.14	self	all features are used, regularized model, C=1000
2	Olga Dalton	85.4	0.06	self, scipy optimize	standardized features, all features are used, Newton-conjugate-gradient optimization
3	Hendrik Maarand	85.4	0.12	self, scikit cross val	all features are used, squared features
4	Margo Kopli	75.0	0.13	self	features 1, 4, 5 and 6

Naive Bayes

Rank	Name	Score	Stdev	Impl	Model details
1	Olga Dalton	90.2	0.07	self	weigh spam emails more heavily by a factor of 1.55
2	Margo Kopli	87.8	0.06	self	smoothing constant $\alpha = 0.01$
3	Ottokar Tilk	87.8	0.07	self	smoothing constant $\alpha = 0.0001$
4	Hendrik Maarand	87.7	0.07	self	plus-one smoothing

Support vector machines

Rank	Name	Score	Stdev	Impl	Model details
1	Ottokar Tilk	85.6	0.00	scikit	sigmoid kernel, regularization $C = 100$
2	Olga Dalton	84.3	0.64	scikit	linear kernel, regularization $C = 0.22$ use a subset of 9 features only
3	Margo Kopli	83.5	0.81	libsvm	linear kernel
4	Hendrik Maarand	83.3	0.88	scikit	linear kernel