Machine Learning - MIRI Master (Final quiz - June 8, 2015)

Tick **clearly** the claims that you think are **true** - deliver just **this sheet** back

Name:

1. **Complexity control and all that.**

1. False
2. False
3. True
4. False - only complex model
5. False

- ❏ If we keep complexity low, we do not need to care about training error
- ❏ Training error is always smaller (or equal) than test error
- ❏ Supplying more training data reduces the chances to obtain an overfitted model
- ❏ Regularization penalizes models that are either simpler or more complex than needed
- ❏ Training error is enough to perform model selection
- ❏ The VC dimension for a two-class classifier penalizes training sample size
- ❏ The VC dimension for a two-class classifier is the maximum number of linear separations that the classifier can perform
- ❏ In order to check that the VC dimension is (at least) some integer k, we just need to find k points that can be shattered
- ❏ Checking that the VC dimension is infinite requires an infinite number of checks
- ❏ A two-class classifier with infinite VC dimension must have an infinite (or very large) number of parameters

2. **Bayesian classifiers.**

- ❏ The Bayes formula converts prior distributions into posterior distributions
- ❏ The Bayes formula is of theoretical importance, but can never be used in practice
- ❏ The numerator in Bayes formula is enough to perform classification
- ❏ The Bayes classifier is the best possible classifier when the prior and posterior distributions are known
- ❏ For normally distributed classes, Bayesian classifiers turn out to be quadratic discriminant functions (QDA)
- ❏ For normally distributed classes, statistical independence among all variables yields linear discriminant functions (LDA)
- ❏ For normally distributed classes, Bayesian classifiers are minimum-distance classifiers
- ❏ The Naive-Bayes classifier assumes statistical independence among all variables
- ❏ The kNN classifier can be explained as a Bayesian classifier
- ❏ The kNN classifier works better with more neighbours, although it is computationally more costly

false

1. True
2. False
3. True
4. True
5. True
6. False
7. True
8. True
9. False
10. False

3. **Maximum Likelihood and GLMs.**

1. True
2. False
3. False
4. False
5.
6. False
7.
8.
9. False
10. False

- ❏ The likelihood of a sample is its density for a given choice of parameters
- ❏ The likelihood of a sample is a function of the sample
- ❏ Logistic regression is a generative linear classifier
- ❏ Logistic regression assumes normally distributed classes
- ❏ In a Generalized Linear Model, the prediction is the logistic function applied to a linear function of the predictors
- ❏ The solution for Logistic regression can be found analytically by minimizing the log-likelihood
- ❏ In Poisson regression, we are interested in predicting integer outcomes, which are equally likely
- ❏ In a Generalized Linear Model, we always find the logistic function in one way or another (so it is called the link function)
- ❏ In statistics, bias and variance are opposite concepts: increasing one must decrease the other

❑ Variance always decreases with increasing sample size; however, bias can increase or stay the same

4. **Regression, linear and non-linear.**

❑ The regression function is the best possible model in regression, and achieves zero error on the training data

❑ The risk is equal to the sum of the (squared) bias, the variance and the noise variance

❑ The theoretical MSE does not depend on the regression function explanation in notion

❑ Models that are "more complex than needed" will tend to have a large bias and large variance

❑ Models that are "less complex than needed" will tend to have a small bias and small variance

❑ A linear combination of non-linear (fixed) functions of the inputs make a linear model

❑ Ridge regression adds a penalty term to a linear model such that the new model is non-linear

❑ Non-linear functions of the data can be estimated by using linear fitting techniques

❑ Both RBFs and MLPs create non-linear models by learning adaptive regressors (regressors with parameters)

❑ Regularization allows the specification of models that are more complex than needed; it also helps numerically

1. False
2. True
3. False
4. False
5. False
6. False
7. False
8. True
9. True
10. True

5. **Kernels and SVMs.**

❑ The kernel function defines kernel matrices whose elements are always positive

❑ The kernel function defines kernel matrices whose elements are always non-negative

❑ The product of two kernel functions may not be a kernel; however, the sum of two kernel functions is always a kernel

❑ By choosing a valid kernel, we get an inner product in a Hilbert space

❑ In order to kernelize a learning algorithm, this must be a supervised one, like the SVM

❑ The cost parameter (C) in a SVM acts as a regularizer of the solution

❑ Increasing the margin in a SVM leads to larger VC dimension

❑ Increasing the margin in a SVM leads to greater chances to separate the data

❑ Increasing the value of C in a SVM, the margin of the solution may increase margin and c have inverse relation

❑ Increasing the value of C in a SVM, the number of training errors may increase

c increase, margin decrease, less misclassification allowed

1. False
2. False - only eigen value non negative, elemnet can be +ve or _ve
3. False - both kernel
4. True
5. Fasle - can be applied to both supervised, unsupervised, ex PCA
6. True
7. True
8.
9. False
10. False

6. **Miscellaneous.**

❑ The k-means algorithm converges to a global optimum if the number of iterations goes to infinity

❑ A Gaussian mixture model assumes normality of the training data individual is gaussian, but whole training data is not gaussian

❑ The k-means algorithm is used to fine-tune a Gaussian mixture model after the latter has converged

❑ The backpropagation algorithm computes the partial derivatives of the error function with respect to the network weights

❑ The backpropagation algorithm must be coupled with an optimization method (update rule) to make it a learning algorithm for a MLP

❑ A MLP requires the specification of the number of hidden neurons; this is best done by trial-and-error, monitoring the fitting error of the network

❑ RBF neural networks are a particular case of MLP networks only activation function might be different but is particular case

❑ In a RBF neural network, regularization does not make sense, because it is based on Euclidean distance regularization makes sense, is based on Euclidean distance

❑ Feature selection can never increase the practical performance of a learning method, it only reduces learning time

❑ Feature selection can be performed after feature extraction, using the extracted features as new variables for selection

1. False
2. False
3. False
4. True
5. True
6. True
7. True
8.
9. False
10. True