# Curse of Dimensionality

1. Which of the following statements describe consequences of the curse of dimensionality?
   1. As dimensionality increases, the volume of the feature space grows exponentially, making data points appear sparse (TRUE)
   2. In high dimensions, most data points tend to be close to the mean of the distribution (FALSE)
   3. In high dimensions, distance metrics such as Euclidean distance become less meaningful (TRUE)
   4. The number of observations required for accurate density estimation increases exponentially with the number of dimensions (TRUE)
2. Indicate which statements characterise the curse of dimensionality:
   1. As the dimensionality increases, data points tend to be all close to each other (FALSE)
   2. As the dimensionality increases, the number of data points in a fixed neighbourhood of an observation quadratically decreases (FALSE)
   3. If the dissimilarity measure is the Euclidean distance, as the dimensionality increases all observations tend to be equally dissimilar (TRUE)
   4. As the dimensionality increases, data points tend to be all far from each other (TRUE)
3. In the context of high-dimensional data, which of the following statements characterize the curse of dimensionality?
   1. Local methods become less effective as dimension increases (TRUE)
   2. The volume of the space increases exponentially with the number of dimensions (TRUE)
   3. High-dimensional spaces require exponentially more sample data to estimate accurately their population density. (TRUE)
   4. In high dimensions, PCA becomes irrelevant (FALSE)

# MVN (Multivariate Normal Distribution)

1. Suppose X follows a multivariate normal distribution with covariance matrix *\Sigma*. Which of the following statements are correct?
   1. In some cases, it is possible to find a linear transformation of X that does not result in another multivariate normal distribution (FALSE)
   2. If X is normally distributed, all of its marginal distributions are also normal (TRUE)
   3. If the components of X are uncorrelated, then they are independent (TRUE)
   4. The determinant of \Sigma is greater than or equal to zero (TRUE)
2. Let X be a random vector with values in R^p. Indicate the true statements:
   1. If X is multivariate Gaussian, the marginal distributions of X are univariate Gaussian (TRUE)
   2. If X is multivariate Gaussian, any linear combination of the coordinates of X is univariate Gaussian (TRUE)
   3. X is multivariate Gaussian if and only if the marginal distributions of X are univariate Gaussian (FALSE)
   4. If one can find a linear combination of the coordinates of X that is univariate Gaussian, then X is multivariate Gaussian (FALSE)
3. Indicate the true statements:
   1. Two random variables whose joint distribution is Gaussian are independent if and only if they are uncorrelated (TRUE)
   2. If two random variables whose joint distribution is Gaussian are uncorrelated then they are independent (TRUE)
   3. If two random variables are uncorrelated then they are independent (FALSE)
   4. If two random variables are dependent then they are correlated (FALSE)
4. Let X be a p-dimensional random vector, following a multivariate normal distribution. Which statements are true?)
   1. If all pairwise correlations between the components of X are zero, they are independent (TRUE)
   2. Any linear combination of the components of X is normally distributed
   3. The marginal distributions of X are univariate normal (TRUE)
   4. The joint distribution of X is completely characterized by its mean vector and covariance matrix (TRUE)
5. Let X\_1, ..., X\_n be an iid sample from a p-dimensional multivariate normal distribution. Which of the following statements are correct?
   1. The squared Mahalanobis distances of the observations from the true mean follow a chi-square distribution (TRUE)
   2. If n is greater than p, then the sample covariance matrix is certainly singular. (FALSE)
   3. The total variance of the sample is the product of the squared Euclidean distances from each observation to the sample mean (FALSE)
   4. The squared Mahalanobis distance of the sample mean from the true mean follows a chi-square distribution. (TRUE)
6. Let X\_1, ..., X\_n be an iid sample from a p-dimensional multivariate Gaussian distribution, with n larger thaan p. Which of the following statements about the Wishart distribution are true?
   1. The sample covariance matrix follows a Wishart distribution
   2. If p=1, the Wishart distribution is equivalent to the a scaled chi-square distribution. (TRUE)
   3. The Wishart distribution is always defined, even when the covariance matrix is singular (FALSE)
   4. The sum of two independent random matrices, each with a Wishart distributions with the same covariance matrix as parameter, has Wishart distribution (TRUE)

# Linear Models

1. Which of the following methods help prevent overfitting in statistical models?
   1. Adding more predictors to the model (FALSE)
   2. Using cross-validation to select model parameters (TRUE)
   3. Applying regularization techniques like ridge regression and LASSO (TRUE)
   4. Directly use the log-likelihood instead of the Akaike Information Criterion (AIC) to perform model selection (FALSE)
2. Indicate which assumptions are consistent with Ordinary Least Squares (OLS):
   1. The data matrix is full-rank (TRUE)
   2. The residuals are serially uncorrelated (TRUE)
   3. The residuals are homoscedastic (TRUE)
   4. The covariates are pairwise independent (FALSE)
3. Cook's distance is used:
   1. In determining if the overall regression model is significant. (FALSE)
   2. As a correction of the Euclidean distance in the features space to account for correlation between the variables (FALSE)
   3. In determining if there is significant collinearity. (FALSE)
   4. In determining the significance of an independent variable. (FALSE)
   5. In identifying influential observations in multiple regression analysis. (TRUE)
4. You are given the reported model. *y = exp(betta0) x1^(betta1)\*x2^(betta2)\*exp(eps)*. Indicate the true statement(s):
   1. This model can be made linear by an exponential transformation of the response variable. (FALSE)
   2. This model can be made linear by a logarithmic transformation of the response variable. (TRUE)
   3. You cannot use the Im() command in R to fit this model. (FALSE)
   4. This model can be made linear by a square root transformation of the response variable. (FALSE)
5. Running a LASSO regression:
   1. Does not require any hyperparameter tuning (FALSE)
   2. Assumes a linear model (TRUE)
   3. Requires the normality assumption (FALSE)
   4. Is a non-parametric regression method (FALSE)
6. Considering the linear model below, where the error term is Gaussian, which of the following hypotheses can be tested with an F-test?

y = Bo + B1x1 + €

* 1. Bo = 0 and B1 = 0 (TRUE)
  2. Bo \* B1 = 1 (FALSE)
  3. Bo = 1 (TRUE)
  4. Bo = 0 (TRUE)

1. Suppose that you fit a linear regression model and the test for the significance of the intercept does not reject the null hypothesis. What are the appropriate conclusions?
   1. There is no statistical evidence to state that the intercept is equal to zero (FALSE)
   2. There is no such test of hypothesis for the intercept (FALSE)
   3. The estimate of the intercept is equal to zero (FALSE)
   4. There is no statistical evidence to state that the intercept is different from zero (TRUE)
2. What is intended by heteroscedasticity?
   1. Residuals do not have zero mean (FALSE)
   2. Residuals are serially correlated (FALSE)
   3. Residuals do not have equal variance (TRUE)
   4. Residuals are not normally distributed (FALSE)
3. The (unbiased) sample covariance matrix of a multivariate Gaussian sample:
   1. Follows a Wishart distribution (TRUE)
   2. Is necessarily invertible (FALSE)
   3. Carries the total variance of the sample that is equal to the sum of its eigenvalues (TRUE)
   4. Is the maximum likelihood estimator of the covariance matrix (FALSE)
4. In a regression problem, you realise that the leave-one-out cross-validation error of your model equals the error computed on the data used to fit the model. Which conclusions can you draw?
   1. Your model has low bias (FALSE)
   2. Your model has low variance (TRUE)
   3. You should go for a simpler model (FALSE)
   4. You are overfitting the data (FALSE)
5. In the context of linear regression, LASSO models are used to:
   1. Incorporate non-linear effects (FALSE)
   2. Avoid overfitting (TRUE)
   3. Perform model selection (TRUE)
   4. Deal with collinearity (TRUE)
6. Consider two linear regression models, M1 and M2, both fitted with OLS on the same dataset and both featuring the intercept along with two continuous regressors which, however, are different in the two models. The likelihood computed for model M1 is 0.01 while for model M2, it is 0.04. Indicate the true statement(s):
   1. In terms of AIC, the best model is M2 (TRUE)
   2. In this case, AIC and BIC drive to different conclusions (FALSE)
   3. In terms of AIC, the best model is M1 (FALSE)
   4. It is not possible to know which model is the best in terms of BIC because some parameters values are missing (FALSE)
   5. In this case, AIC and BIC drive to the same conclusion (TRUE)
7. Which of the following observations are signs of a collinearity issue in a linear regression model?
   1. High variance inflation factor for one or more variable (TRUE)
   2. Low variance inflation factor for one or more variable (FALSE)
   3. Flat ellipsoidal confidence region for the vector of estimated coefficients (TRUE)
   4. Excessively large confidence interval for one or more estimated coefficients (TRUE)
8. Which of the following statements are true in high-dimensional regression models?
   1. Ridge regression is particularly appropriate when we expect only a few predictors to have a significant effect on the target variable (FALSE)
   2. Ridge regression shrinks coefficients towards zero to reduce their variability (TRUE)
   3. LASSO regression shrinks coefficients towards zero to reduce their variability (TRUE)
   4. Increasing the number of predictors will never decrease the coefficient of determination R^2 (TRUE)

# R^2

1. In a linear model, the coefficient of determination:
   1. It's a metric to perform model selection between models of similar complexity (TRUE)
   2. Can take values between -1 and 1 (FALSE)
   3. Allows to assess the fit of the model to the data (TRUE)
   4. Indicates perfect interpolation if it is equal to 1 (TRUE)
2. Suppose that the value of R^2 for an estimated regression model is exactly zero. Which of the following are true?
   1. The intercept coefficient estimate must be zero. (FALSE)
   2. The regression line has not explained any of the variability of target variable around its mean value (TRUE)
   3. But for the intercept, all the coefficient estimates are zero (TRUE)
   4. The fitted model will be an horizontal line with respect to all of the explanatory variables (TRUE)

# Multiple Hypothesis testing

1. Which of the following statements about multiple hypothesis testing are correct?
   1. The Bonferroni correction controls the Familywise Error Rate (FWER) (TRUE)
   2. The False Discovery Rate (FDR) is the expected proportion of false positives among rejected hypotheses (TRUE)
   3. The Bonferroni correction is less conservative than the Benjamini-Hochberg procedure (FALSE)
   4. The Bonferroni correction controls the False Discovery Rate (FDR) (FALSE)
2. You are running a large-scale hypothesis testing procedure that guarantees you a False Discovery Rate (FDR) at level 10%. Indicate the true statements:
   1. The Family Wise Error Rate is larger than or equal to 10%. (TRUE)
   2. The expected proportion of true H0s that will be rejected over the total of rejected hypothesis is 10% (setting by convention the proportion to zero if there are no rejections) (TRUE)
   3. If the tests are not independent, one can control the FDR at level 10% with the Bonferroni strategy (FALSE)
   4. If the tests are independent, one can control the FDR at level 10% with the Benjamini & Hochberg strategy (TRUE)
3. We are considering confidence intervals for linear combinations of the mean. Indicate the true statements:
   1. One-at-a-time t2 intervals cannot be wider than Bonferroni confidence intervals (TRUE)
   2. Bonferroni intervals are always narrower than simultaneous T2 intervals (FALSE)
   3. Bonferroni intervals are always wider than simultaneous T2 intervals (FALSE)
4. Consider a random variable X and the two random intervals I and J defined below. Indicate the true statements.

1) I is a 95% confidence interval for *mu* = E(X)

2) J is a 95% prediction interval for X

* 1. P( *mu* € *I* ) < 0.95 (FALSE)
  2. I will contain *mu* with probability equal to or larger than 95%
  3. P(X € J) = 0.95 (TRUE)

1. How does the Bonferroni correction adjust the significance level for multiple comparisons? Indicate the true statement(s):
   1. It adds a constant value to the original significance level (FALSE)
   2. It multiplies the original significance level by the number of comparisons (FALSE)
   3. It does not change the significance level (FALSE)
   4. It divides the original significance level by the number of comparisons (TRUE)
2. Which of the following statements about the False Discovery Rate (FDR) and the Family-Wise Error Rate (FEWER) are true?
   1. The Bonferroni method with a global significance level of 5% ensures the expected FDR is less than or equal to 5% (TRUE)
   2. When there are no discoveries to be made, FWER control is preferable to FDR control (FALSE)
   3. The Benjamini-Hochberg procedure controls the FWER under independence (FALSE)
   4. The Bonferroni method with a global significance level of 5% ensures the FWER is less than or equal to 5% (TRUE)
3. You perform simultaneous inference on a p-dimensional multivariate mean vector of a Gaussian distribution, with p larger than or equal to 2, using Bonferroni-corrected confidence intervals. Which statements are true?
   1. The Bonferroni correction controls the FWER at the specified level (TRUE)
   2. Simultaneous Hotelling's T^2 intervals are always wider than Bonferroni intervals (FALSE)
   3. Bonferroni corrected t-intervals are always wider than the corresponding one-at-a-time t-intervals (TRUE)
   4. The Bonferroni method cannot be applied if the p-components are not independent (FALSE)

# Covariance Matrix

1. Let S be the sample covariance matrix of n observations of a p-dimensional dataset. Which of the following statements are necessarily true?
   1. S is always invertible if n > p (FALSE)
   2. If S has at least one zero eigenvalue, then the dataset has linearly dependent variables (TRUE)
   3. The determinant of S represents the total variance of the dataset (FALSE)
   4. If all pairwise sample correlations are zero, then S is a diagonal matrix (TRUE)
2. 7. Let X\_1, …, X\_n be a sample from a multivariate distribution. Which of the following statements about the sample covariance matrix S are true?
   1. S always admits an eigendecomposition (TRUE)
   2. tr(S) represents the generalized variance (FALSE)
   3. The Mahalanobis distance accounts for the correlation structure in the data (TRUE)
   4. S is positive definite if the data matrix is full rank (TRUE)

# Mahalanobis distance

1. Which of the following statements are true about the Mahalanobis distance?
   1. It's not a metric on R^p (FALSE)
   2. It does not account for correlations among variables (FALSE)
   3. If the covariance matrix is the identity matrix, it coincides with the Euclidean distance (TRUE)
   4. It is invariant under orthogonal transformations of the data (TRUE)

# Multidimensional scaling (MDS)

1. Which of the following statements about classical multidimensional scaling (MDS) are true?
   1. MDS looks for a Euclidean representation of observations based on their pairwise distance matrix (TRUE)
   2. The input to MDS must be a Euclidean distance matrix (FALSE)
   3. MDS is equivalent to PCA when applied to standardized data (FALSE)
   4. MDS tries to match the original distances with the Mahalanobis distances in the Euclidean representation space (FALSE)

# Hotelling’s T^2 test

1. Which of the following are true about Hotelling's T2 test?
   1. It generalizes the t-test to the multivariate case (TRUE)
   2. It assumes that the data follows a multivariate normal distribution (TRUE)
   3. Modulo a scaling factor, the T42 statistic follows a Chi-square distribution (FALSE)
   4. When n is large compared to p, it can be used without the assumption of multivariate normality (TRUE)
2. Which statements about Hotelling's T2 test are true?
   1. The test statistic follows an exact F-distribution (up to a scaling factor) under the null hypothesis when n goes to infinity, otherwise it follows an approximate F-distribution only (FALSE)
   2. It is equivalent to a student t-test in the univariate case (TRUE)
   3. The test assumes the covariance matrix is known (FALSE)
   4. The test statistic follows an exact F-distribution (up to a scaling factor) under the null hypothesis (TRUE)

# (M)ANOVA

1. Which of the following differences between MANOVA and ANOVA are correct?
   1. ANOVA compares means of univariate distributions, while MANOVA compares means of multivariate distributions (TRUE)
   2. MANOVA takes into account the dependence between the components of the observation vector, whereas individual ANOVA's, one for each component, do not. (TRUE)
   3. If a MANOVA test is significant, all individual ANOVA tests, one for each component of the observation vector, must also be significant (FALSE)
   4. MANOVA requires that all components of the observation vector have the same variance. (FALSE)
2. One-way ANOVA:
   1. Does not require groups to have the same variance (FALSE)
   2. Is equivalent to estimating the parameters of a well-chosen linear model with ordinary least squares (TRUE)
   3. Requires to have equal group sample sizes (FALSE)
   4. Allows to jointly analyse the effect of two factors on the response (FALSE)
3. You have conducted an ANOVA to test the mathematical ability of four different groups of students. The result of the analysis are summarized in the following table:

| Source of variability | df | Sum of squares |
| --- | --- | --- |
| Between groups | 3 | 625 |
| Within groups | 36 | 2128 |

* 1. The value of the F statistic is 0.29 (FALSE)
  2. With the information provided by the table it is impossible to compute the p-value of the F test (FALSE)
  3. With the information provided by the table it is impossible to know the total number of students tested (FALSE)
  4. The total number of students tested was 39 (FALSE)
  5. The value of the F statistic is 3.52 (TRUE)
  6. The p-value of the F test is lower than 5% (TRUE)

1. Which of the following statements correctly describe differences between MANOVA and ANOVA?
   1. MANOVA requires equal group sizes across treatments (FALSE)
   2. A significant MANOVA implies a significant ANOVA for at least one component (FALSE)
   3. A significant MANOVA implies at least one group mean vector differs (TRUE)
   4. two-ways MANOVA allows for interaction between the treatment effects, while two-ways ANOVA does not (FALSE)

# PCA

1. Suppose you conduct PCA on a dataset with p variables. Which of the following statements are true?
   1. The first principal component explains a proportion of the total variance that is larger than or equal to that of any of the p variables (TRUE)
   2. The first principal component explains a proportion of the total variance that is larger than or equal to that of any of the other principal components (TRUE)
   3. If the p variables have been standardized to have zero mean and unit variance, then they all contribute equally to the total variance (TRUE)
   4. The sum of the eigenvalues of the covariance matrix equals the generalised variance of the dataset. (FALSE)
2. What is the typical range of values for eigenvalues in PCA when analyzing a dataset?
   1. Non negative real numbers multiple of the variance (FALSE)
   2. Integers (FALSE)
   3. Non negative real numbers (TRUE)
   4. Real numbers between 0 and 1 (FALSE)
   5. Real numbers (FALSE)
3. About Principal Component Analysis (PCA):
   1. The total variance is distributed among principal components proportionally to the eigenvalues of the data covariance matrix (TRUE)
   2. The total variance is distributed among principal components proportionally to their number of strictly positive loadings (FALSE)
   3. The eigenvalues of the covariance matrix can never be all equal (FALSE)
   4. If variables are first standardized, the eigenvalues of their sample covariance matrix are all the same. (FALSE)
4. Which of the following statements are true regarding PCA?
   1. PCA provides uncorrelated linear combinations of the original variables. (TRUE)
   2. Standardizing variables before PCA ensures that all eigenvalues are equal (FALSE)
   3. The total variance explained by the first k principal components equals the sum of the k largest eigenvalues of the sample covariance matrix. (TRUE)
   4. The eigenvectors of the sample covariance matrix form an orthonormal basis. (TRUE)
5. Which statements are correct regarding scaling before PCA?
   1. Without scaling, variables with large variance dominate the first PCs (TRUE)
   2. Centering without scaling is appropriate if variables are on the same scale (TRUE)
   3. Standardizing is equivalent to applying the PCA to the correlation matrix instead of the covariance matrix (TRUE)
   4. PCA is unaffected by whether data are scaled or not (FALSE)
6. Suppose you perform PCA on standardized 10-dimensional multivariate data. Which of the following statements are true?
   1. The first principal component corresponds to the direction of maximum correlation (FALSE)
   2. The loadings matrix is symmetrical (FALSE)
   3. The scores are linear combinations of the original variables. (TRUE)
   4. The total variance is equal to 1 (FALSE)

# LDA/QDA/FDA

1. Which of the following properties hold for Linear Discriminant Analysis (LDA)?
   1. If the two classes are normally distributed but have different covariance matrices, LDA is the optimal classifier (FALSE)
   2. If the two classes have identical covariance matrices but are not normally distributed, LDA is the optimal classifier (FALSE)
   3. LDA assumes that the within-class covariance matrices are equal for all classes (TRUE)
   4. If the two classes are normally distributed and have identical covariance matrices, LDA is the optimal classifier (TRUE)
2. Fisher Discriminant Analysis
   1. Does not require the assumption of normality (TRUE)
   2. Is a particular case of QDA (FALSE)
   3. Generates a classifier different from a Bayes classifier (FALSE)
   4. Is a particular case of LDA (TRUE)
3. In Principal Component Analysis (PCA), the loadings:
   1. Are the projection of the observations on the directions of the principal components (FALSE)
   2. Are necessarily positive (FALSE)
   3. Are the coefficients of the change-of-basis matrix P (TRUE)
   4. Always sum to 1 (FALSE)
4. Suppose we are working on a binary classification problem with equal prior probabilities. Suppose that observations are normally distributed in each class with same covariance matrix and that the costs of misclassifications are c(1|0) = 1 and c(0|1) = 2. Indicate the true statements:
   1. A Linear Discriminant Analysis with prior probabilities p(0)=1/3 and p(1)=2/3 minimises the expected cost of misclassification (TRUE)
   2. A Linear Discriminant Analysis with equal prior probabilities minimises the Actual Error Rate (TRUE)
   3. A Fisher Discriminant Analysis minimises the Actual Error Rate (TRUE)
   4. A Linear Discriminant Analysis with prior probabilities p(0)=2/3 and p(1)=1/3 minimises the expected cost of misclassification (FALSE)
5. Which of the following statements are correct regarding Linear Discriminant Analysis (LDA)?
   1. LDA is always mathematically equivalent to Fisher's discriminant analysis (FALSE)
   2. LDA assumes equal covariance matrices across classes (TRUE)
   3. LDA maximizes within-class variability while minimizing between-class variability (FALSE)
   4. LDA is robust to violations of the normality assumption (TRUE)
6. Suppose you have two groups of sampled data, both generated by a p-dimensional multivariate Gaussian distribution, with the two distributions having the same covariance matrix. Which statements are true in this case?
   1. LDA is preferable to QDA (TRUE)
   2. Fisher discriminant analysis is preferable to LDA (FALSE)
   3. LDA is not to be used if the samples are unbalanced (FALSE)
   4. LDA is an instance of Bayes classifier (TRUE)

# CLUSTERING

1. About DBSCAN
   1. DBSCAN can leave some points uncluttered as noise (TRUE)
   2. DBSCAN is suited for identifying clusters of varying densities (FALSE)
   3. DBSCAN is suited for identifying non-convex clusters (TRUE)
   4. Scaling or not the variables has no effect on DBSCAN (FALSE)
2. You are given a dissimilarity matrix between N statistical units. Which of the following methods сould you apply to cluster the units?
   1. k-nearest neighbours (FALSE)
   2. Hierarchical clustering with Ward linkage (FALSE)
   3. K-means (FALSE)
   4. DBSCAN (TRUE)
   5. Hierarchical clustering with average linkage (TRUE)
3. The cophenetic distance:
   1. Is commonly used to choose between hierarchical clustering, DBSCAN and k-means methods for clustering. (FALSE)
   2. Is a metric (TRUE)
   3. Is an ultrametric (TRUE)
   4. In hierarchical clustering, is the height of the dendrogram where the two branches that include two observations merge into a single branch (TRUE)
4. Which statements about DBSCAN are correct?
   1. DBSCAN requires the number of clusters as an input parameter (FALSE)
   2. DBSCAN is sensitive to the choice of distance metric and scale of variables. (TRUE)
   3. Border points do not belong to any cluster (FALSE)
   4. DBSCAN is suitable for convex clusters only (FALSE)
5. You are given a dissimilarity matrix D. Which of the following clustering methods can you apply directly without accessing the original feature space?
   1. DBSCAN (TRUE)
   2. Hierarchical clustering with Ward linkage (FALSE)
   3. K-means (FALSE)
   4. Hierarchical clustering with complete linkage (TRUE)

# Classifier Metrics

1. You built a binary classifier; its confusion matrix on the test set is presented below. The prior probabilities are given by the observed relative frequencies. The costs of misclassification are c(yes|no) = 5 and c(no|yes) = 1
   1. The expected cost of misclassification is 1/3 (TRUE)
   2. The estimated probability that a new observation classified as "yes" is in fact a "yes" is 0.60 (rounded to two decimal places) (FALSE)
   3. The estimated probability that a new observation is correctly classified as "yes" is 0.60 (rounded to two decimal places) (TRUE)
   4. The estimated probability that the classifier predicts a new observation as "no" is ⅓ (TRUE)

| n=165 | **Predicted: NO** | **Predicted: YES** |
| --- | --- | --- |
| **Actual: NO** | 50 | 10 |
| **Actual: YES** | 5 | 100 |

1. You built a binary classifier; its confusion matrix on a test set is presented below. The prior probabilities are given by the observed relative frequencies. The costs of misclassification are both equal to 1.
   1. The estimated prior probability of the positive class is 0.67 (rounded to 2 decimal places) (FALSE)
   2. The estimate of the Actual Error Rate is 0.09 (rounded to 2 decimal places) (TRUE)
   3. The expected cost of misclassification is 0.09 (rounded to 2 decimal places) (TRUE)
   4. The estimated probability that an observation predicted as "no" is in fact a "yes", is equal to 0.09 (rounded to two decimal places) (TRUE)

| n=165 | **Predicted: NO** | **Predicted: YES** |
| --- | --- | --- |
| **Actual: NO** | 50 | 10 |
| **Actual: YES** | 5 | 100 |

1. smth

# KNN (K-nearest-neighbors)

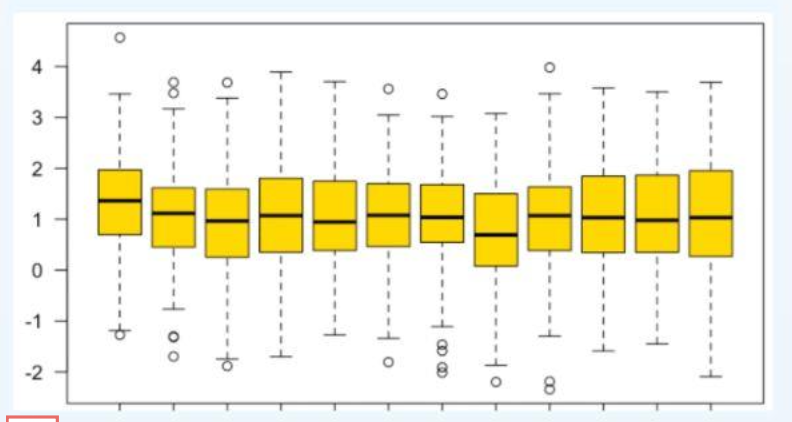
1. When you classify observations with k-nearest neighbours:
   1. You are using a parametric method (FALSE)
   2. Taking k=1 exposes you to a high bias in the estimation (FALSE)
   3. Even in a high-dimensional setting with a small sample size, you will perform well (FALSE)
   4. You don't have to test the normality of the data (TRUE)
2. k-nearest neighbours (k-NN):
   1. Tends to overfit the training data as k gets larger (FALSE)
   2. Is a classification method (TRUE)
   3. Is a clustering method (FALSE)
   4. Tends to overfit the training data as k gets smaller (TRUE)
3. About the k-NN classification method:
   1. It assumes multivariate normal observations (FALSE)
   2. The bias increases as the number k of neighbours increases (TRUE)
   3. The variance increases as the number k of neighbours increases (FALSE)
   4. In general, it doesn't work well when the data points dimensionality is high with respect to the sample size (TRUE)

# SVM (Support vector Machines)

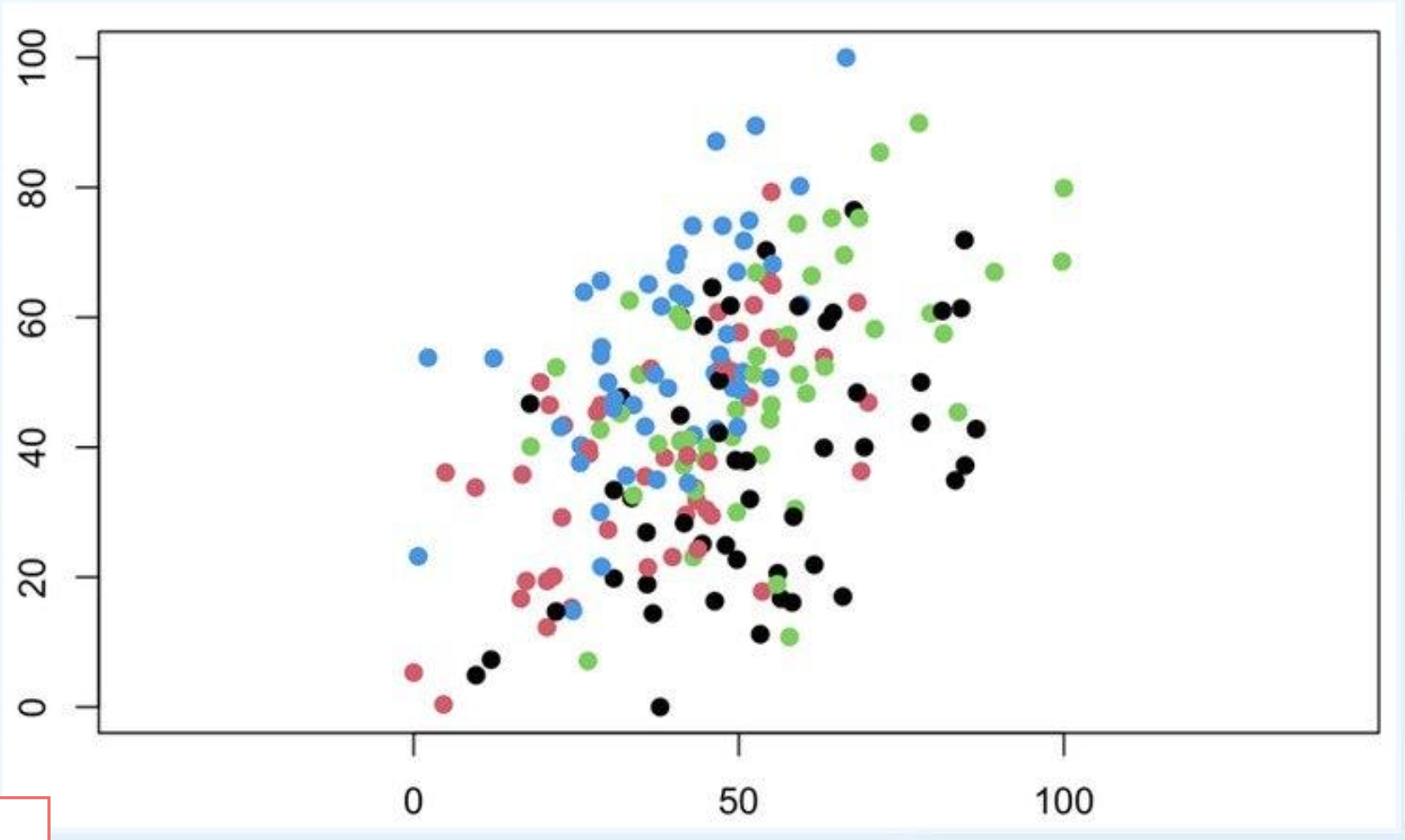
1. In Support Vector Machines:
   1. It is always preferable to use a non-linear kernel (FALSE)
   2. Increasing the budget constraint reduces the variance (TRUE)
   3. There is always the possibility to find a separating hyperplane (FALSE)
   4. We want to find an hyperplane that best separates the two classes (TRUE)

# Other questions

1. Consider a dataset of 100 observations and 12 variables expressed in the same unit, whose marginal distributions are represented by the following box plots. Which statements do you agree with?
   1. The variables have zero sample mean (FALSE)
   2. The variables have equal sample variance (FALSE)
   3. It is necessary to scale the variables before perform in a Principal Component Analysis (FALSE)
   4. The variables may have zero mean (FALSE)
   5. The variables may have equal variance (TRUE)



1. Consider the bivariate observations of four groups, plotted below (colored by group). Indicate the true statements:
   1. The assumption of Gaussianity does not seem unreasonable (TRUE)
   2. The covariance matrix looks different for each group (FALSE)
   3. We can carry on a MANOVA to test for the difference in the mean between the groups (TRUE)
   4. The two variables are correlated (TRUE)



1. The one-parameter Box-Cox transformation: Indicate the true statements:
   1. Allows to solve the problem caused by the outliers (FALSE)
   2. Is based on a power transformation of data (TRUE)
   3. Guarantees multivariate normality when it successfully makes each marginal univariate normal (FALSE)
   4. Always allows to make data normal (FALSE)
   5. Is the only possible transformation to make data normal (FALSE)
   6. Includes the log transformation (TRUE)
2. Consider n independent observations generated by a continuous distribution defined on R^p, with p strictly greater than n.
   1. The rank of the data matrix is strictly less than n, with probability one (FALSE)
   2. The generalised variance is equal to 0 (TRUE)
   3. The total variance is equal to 0 (FALSE)
   4. The rank of the data matrix is equal to n, with probability one
   5. The covariance matrix is singular (TRUE)
   6. The columns of the data matrix are linearly dependent (TRUE)
3. Which statements about the generalized variance of a p-dimensional multivariate sample are true?
   1. It is the product of the eigenvalues of the sample covariance matrix (TRUE)
   2. It is invariant to rotations of the reference system (TRUE)
   3. It increases as correlation between variables increases (FALSE)
   4. It is zero if the sample points lie in a hyperplane of dimension p-1. (TRUE)