Università degli Studi di Padova

Department of Industrial Engineering

Machine learning
Lesson #15

Academic year 2025-2026

Prof. Pierantonio Facco

CAPE-Lab, Computer-Aided Process Engineering Laboratory

Email: pierantonio.facco@unipd.it

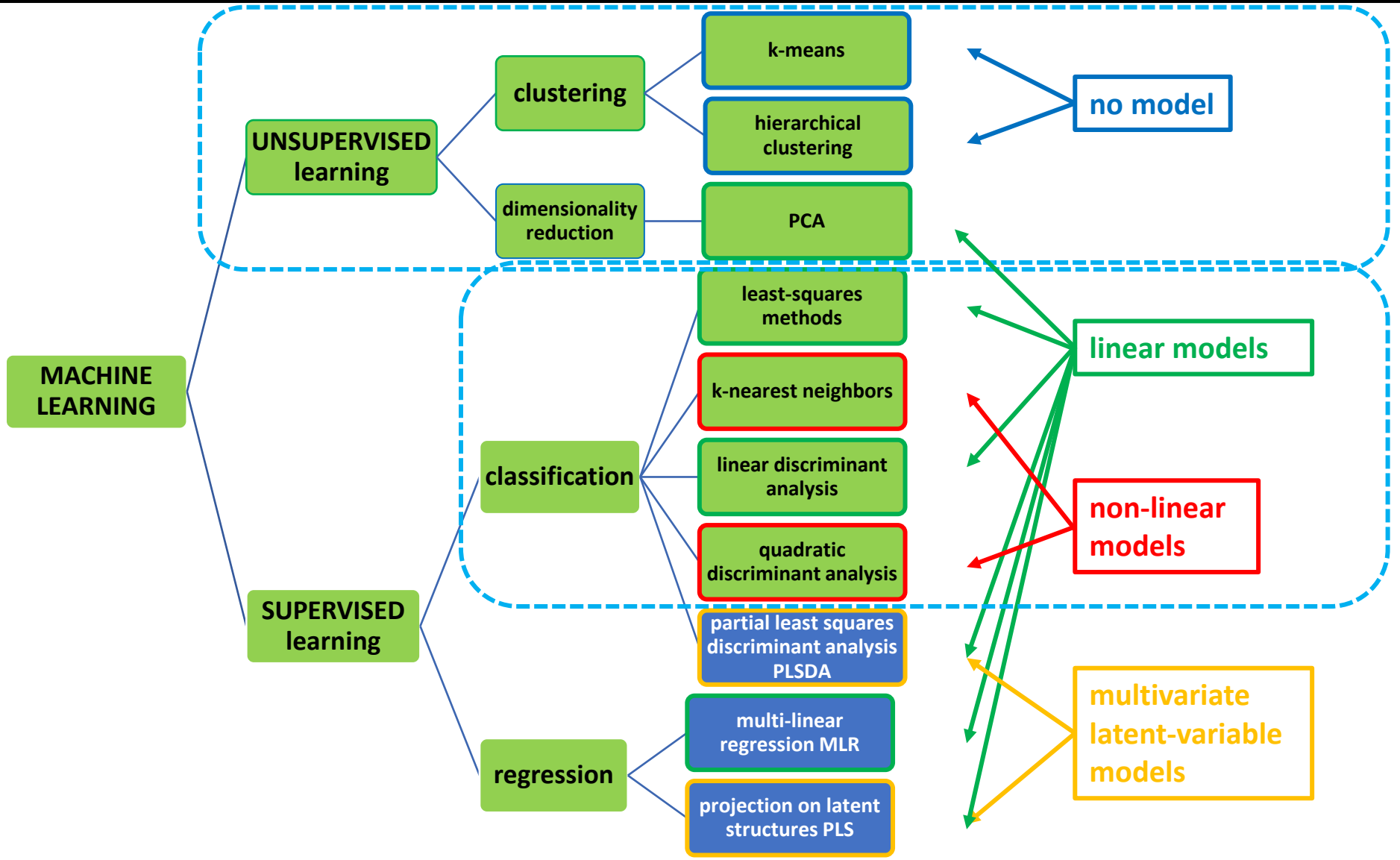
URL: https://research.dii.unipd.it/capelab/

Gentle reminder: next lesson is a flipped one

- Please, complete the following procedure before next lesson:
  1. **attend the video lecture #16**
  2. **self-assess your learning**
  3. **re-read the following papers** available in the "Suggested reading" in the course Moodle, especially for the PLS part:
     - Geladi, P., Kowalski, B.R. (1986). Partial least-squares regression: a tutorial. *Anal. Chim. Acta,* **185**, 1–17
     - Wise, B.M., Gallagher, N.B. (1996). The process chemometrics approach to process monitoring and fault detection. *J. Process Control,* **6**, 329–348
  4. **prepare questions** and anything you need to discuss with the teacher and your mates in the next lecture

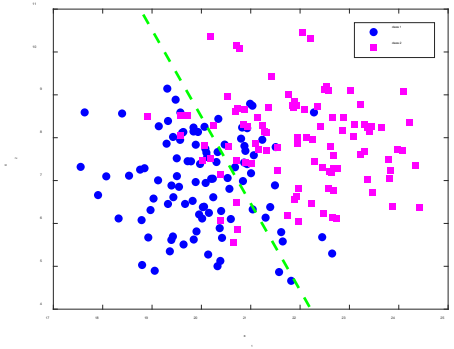
- Next lesson will be held in the following manner:
  - 1/2 Q&A: questions (of the students) and answers (of the teacher)
  - 1/2 we will begin the second part of the course, Design of Experiments

Recap and contextualization



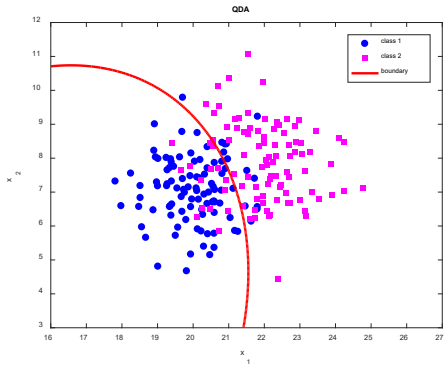
Recap on classification methods

- **Least-squares methods**
  - a linear regression method is used
  - linear decision boundary
  - overlap between classes (misclassifications)
  - appropriate for 2 multi-normal classes

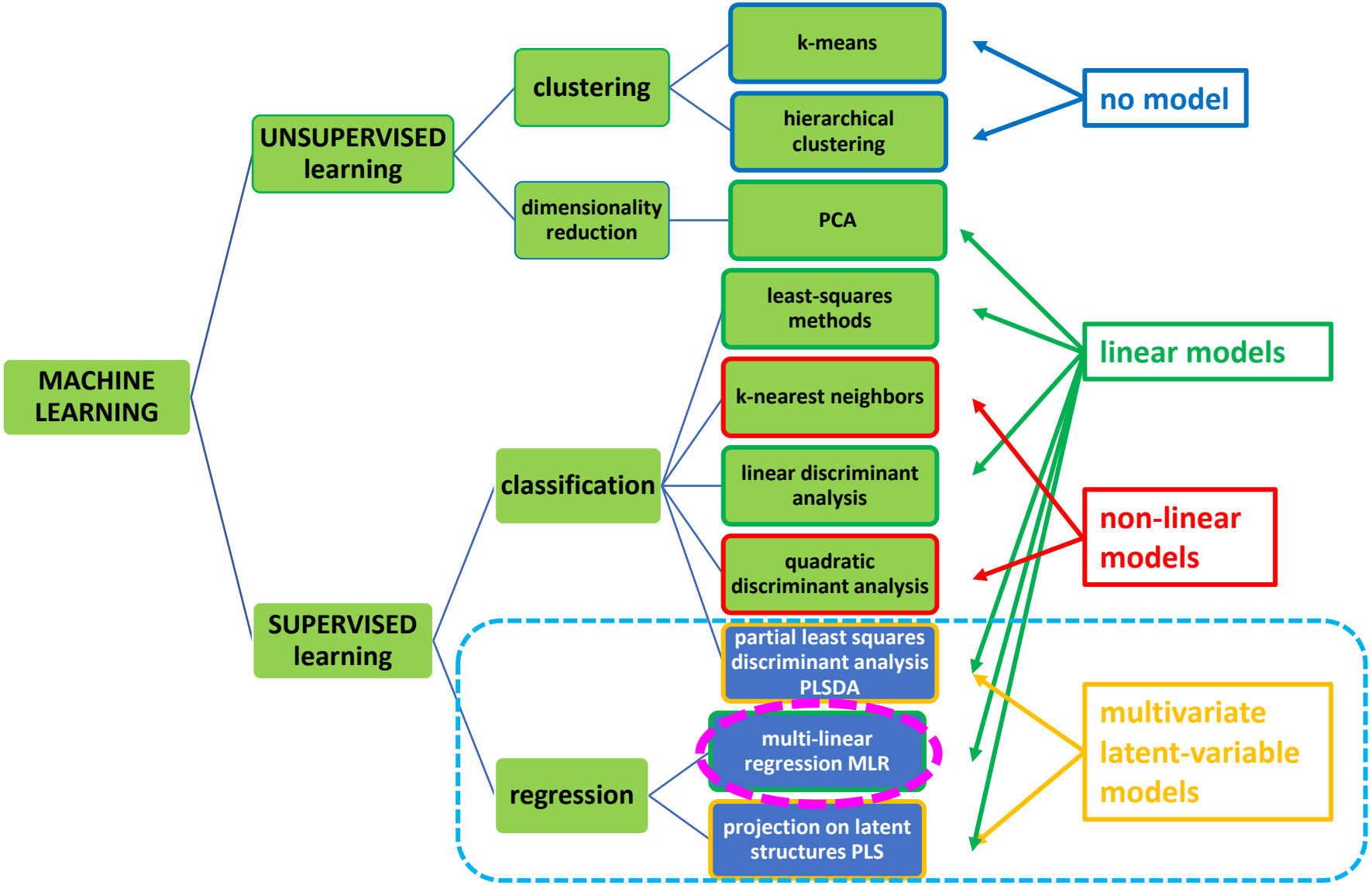


- ***K*-nearest neighbors**
  - no assumptions
    - similarity based on closeness
  - irregular boundary
  - appropriate for mixtures of multi-normal populations

- **Linear and quadratic discriminant analysis**
  - based on a decision rule which is the maximization of a discriminant function
    - maximum similarity within classes
    - maximum dissimilarity between classes
  - appropriate for multivariate distributions with the same covariance structures



Recap and contextualization



Linear regression models

Linear regression models

- In engineering applications, it is typically valuable to understand the relation between two (or more) groups of variables
  - example: to the purpose of process prediction, optimization or control, understanding the relation among:
    - product yield in a process (a continuous variable in the domain of real numbers)
    - raw material properties, process inputs and settings, and process variables
- In most cases the true functional relation among the response (dependent) variable y and the V independent regressors $[x_1, x_2, \dots, x_V]$ is unknown:
  - **simple empirical regression models** are desirable
  - a **low order polynomial** can be effectively utilized to fit the available data
    - widely used as an approximation function
- **Multiple Linear Regression (MLR)** models relate the V predictor variables x_v (i.e., regressors) to the response variables y in a linear function of parameters β_v by means of:

$$y = \beta_0 + \sum_{v=1}^{V} \beta_v x_v + e$$

  - where the β_v are **regression coefficients**
    - represent the expected change in y per unit change in x_v when all the remaining independent variables are held constant
  - the mathematical expression describes a hyperplane in the *V*-dimensional space of the regressors

Higher complexity linear regression models

- Different terms can be included the **linear** model to add complex relations between regressors x_v and response y, such as:
  - **interaction terms**:

$$y = \beta_0 + \sum_{v=1}^{V} \beta_v x_v + \sum_{v_1=1}^{V} \sum_{v_2=1}^{V} \beta_{v_1,v_2} x_{v_1} x_{v_2} + e$$

even if higher order terms are included, **the model is linear** because only linear functions of the coefficients are used

  - **second-order terms**:

$$y = \beta_0 + \sum_{v=1}^{V} \beta_v x_v + \sum_{v_1=1}^{V} \sum_{v_2=1}^{V} \beta_{v_1,v_2} x_{v_1} x_{v_2} + \sum_{v=1}^{V} \beta_{v,v} x_v^2 + e$$

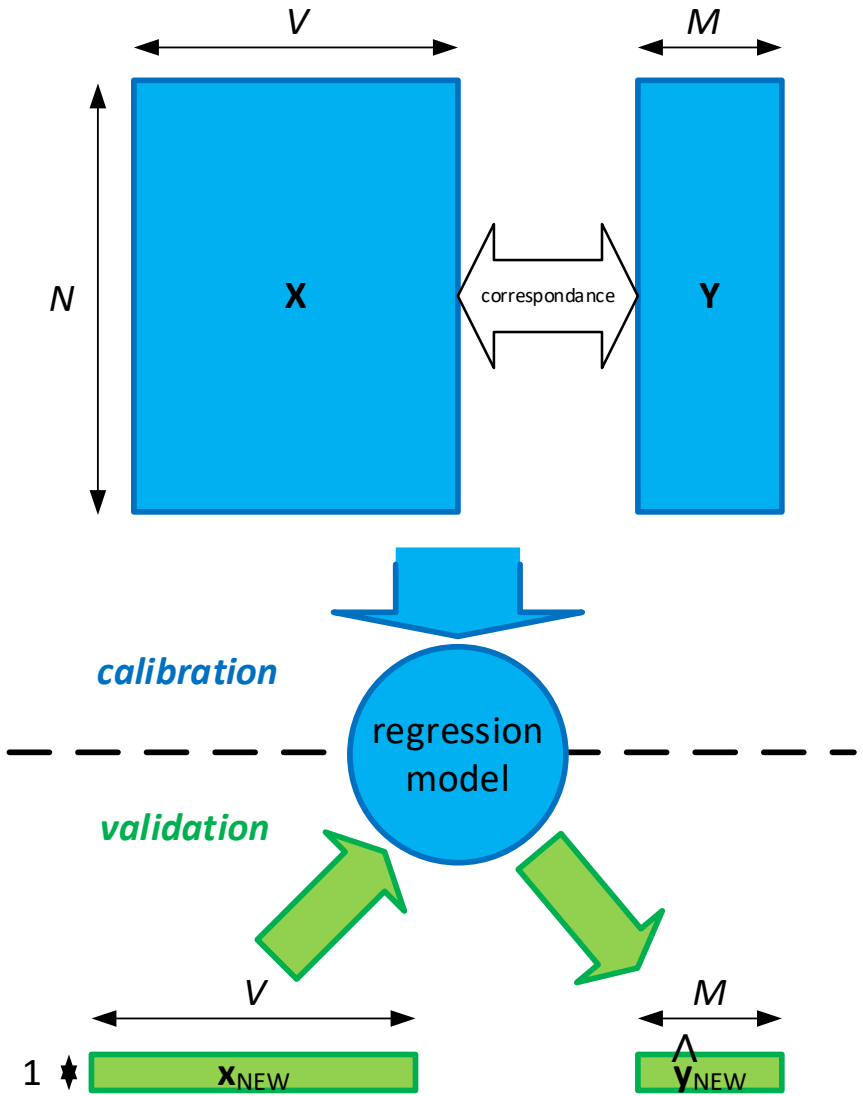
  - furthermore, some *tricks* can be used to succeed in MLR modelling:
    - nonlinear transformations of the original variables (e.g.: logarithms or exponentials of the variables)
    - use of dummy variables
    - etc.

General regression problem formalization

- The N **observations of the** V **measured inputs** $\mathbf{X}$ $[N \times V]$ (i.e., regressors, predictors, independent variables, usually measured at a high frequency) are available
- The **corresponding** N **observations of the** M **output variables** $\mathbf{Y}$ $[N \times M]$ (i.e., regressed variables, estimated or predicted variables, dependent variables, usually measured at a low frequency) are available, as well

- Regression parameters $\hat{\boldsymbol{\beta}}$ are estimated from the available data
- The **estimated regression parameters** $\hat{\boldsymbol{\beta}}$ are used to **estimate/predict the response variable(s)** $\hat{y}_{\mathrm{NEW}}$ for new observations where the V predictors $\mathbf{x}_{\mathrm{NEW}}$ are available

Schematic of the regression problem formalization



available $\mathbf{X}$ and $\mathbf{Y}$ data are used to calibrate the regression model, namely, to estimate the regression coefficients $\widehat{\mathbf{B}}$

the regression coefficient estimated by means of the calibration data are then used to perform predictions on new incoming data

Discussion

- What could be a good strategy to estimate the model parameters?

Parameter estimation

- When both regressors and responses are available one can build a regression model
  - **a linear regression model can be trained from the available data by estimating the parameters β_v**
  - the best strategy to fit the model is **minimizing the error**
- The β_v **parameters estimation** is usually carried out by means of a **least-squares method**:
  - consider the case of observation $n = 1, 2, \dots, N$:

$$y_n = \beta_0 + \sum_{v=1}^{V} \beta_v x_{n,v} + e_n$$

  - the least-squares method calculates the $V + 1$ β_v parameters in such a way as to **minimize the sum of the squares of the errors** e_n:

$$RSS = \sum_{n=1}^{N} e_n^2 = \sum_{n=1}^{N} \left(y_n - \beta_0 - \sum_{v=1}^{V} \beta_v x_{n,v} \right)^2$$

  - this function is to be minimized with respect to the regression coefficients to satisfy:

$$\left. \frac{\partial RSS}{\partial \beta_v} \right|_{\beta_0, \beta_1, \dots, \beta_V} = -2 \sum_{n=1}^{N} \left(y_n - \beta_0 - \sum_{v=1}^{V} \beta_v x_{n,v} \right) x_{n,v} = 0 \quad \forall\, v = 1, 2, \dots, V$$

Least-squares solution

- The estimated regression coefficients in *vector/matrix form* are:

$$RSS = \mathbf{e}^{\mathrm{T}}\mathbf{e} = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^{\mathrm{T}}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$$

$$\left.\frac{\partial RSS}{\partial \boldsymbol{\beta}}\right|_{\hat{\boldsymbol{\beta}}} = -2\mathbf{X}^{\mathrm{T}}\mathbf{y} + 2\mathbf{X}^{\mathrm{T}}\mathbf{X}\hat{\boldsymbol{\beta}} = 0$$

$$\hat{\boldsymbol{\beta}} = \left(\mathbf{X}^{\mathrm{T}}\mathbf{X}\right)^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{y}$$

- The fitted regression model is:

$$\hat{\mathbf{y}} = \mathbf{X}\hat{\boldsymbol{\beta}}$$

  - and for the n-th observation is:

$$\hat{y}_n = \mathbf{x}_n^{\mathrm{T}}\hat{\boldsymbol{\beta}}$$

  - and $\mathbf{e}$ is the residual vector:

$$\mathbf{e} = \mathbf{y} - \hat{\mathbf{y}}$$

    - where $E(\mathbf{e}) = 0$ and $V(\mathbf{e}) = \sigma^2$

Model verification

- The meaningfulness of the regression model must be verified by answering the following questions to understand if the regression model is appropriate and reliable:
  1. how can the **validity of the model** be verified?
  2. are **all the terms of the regression model meaningful**?
  3. how can the **model uncertainty** be determined?
  4. are the **data appropriate** to build the model?
  5. is the **linear model structure appropriate**?

Discussion

- Could you think what could be a measurement of model adequacy and how inappropriate data can be found?

Model validity

Model validity: coefficients of determination

- The **coefficient of multiple determination** for a variable y:

$$R^2 = \frac{SS_{\mathrm{R}}}{SS_{\mathrm{T}}} = 1 - \frac{SS_{\mathrm{E}}}{SS_{\mathrm{T}}} = 1 - \frac{\sum_{n=1}^{N}(y_n - \hat{y}_n)^2}{\sum_{n=1}^{N}(y_n - \bar{y}_n)^2}$$

  - SS_{R} is the sum of squares of the regression, $SS_{\mathrm{E}} = RSS$, and SS_{T} is the total sum of squares
  - it is a measure of the amount of reduction in the variability of y obtained by using the regression model

- The **adjusted determination coefficient** is:

$$R^2_{\mathrm{adj}} = 1 - \frac{\frac{SS_{\mathrm{E}}}{(N-V)}}{\frac{SS_{\mathrm{T}}}{(N-1)}} = 1 - \frac{(N-1)}{(N-V)}\left(1 - R^2\right)$$

  - it does not always increase as variables are added to the model
    - if unnecessary terms are added, the value of R^2_{adj} will often decrease
    - when R^2 and R^2_{adj} differ dramatically, there is a good chance that **non-significant terms** have been included in the model

in this way one can verify if the model is valid

Residual sum of squares

- The residual sum of squares are:

$$SS_{\mathrm{E}} = \sum_{n=1}^{N} (y_n - \hat{y}_n)^2 = \sum_{n=1}^{N} e_n^2 = \mathbf{e}^{\mathrm{T}}\mathbf{e} = \left(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}\right)^{\mathrm{T}}\left(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}\right)$$

$$SS_{\mathrm{E}} = \mathbf{y}^{\mathrm{T}}\mathbf{y} - \hat{\boldsymbol{\beta}}^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}\mathbf{y}$$

  - they have $(N - V)$ degrees of freedom and $E(SS_{\mathrm{E}}) = (N - V)\sigma^2$

- The **error variance** is:

$$\sigma^2 = \frac{SS_{\mathrm{E}}}{N - V}$$

- The least-squares method produces an unbiased estimator of the regression coefficients (it can be demonstrated), whose **covariance** is:

$$\mathrm{cov}\left(\hat{\boldsymbol{\beta}}\right) = \sigma^2\left(\mathbf{X}^{\mathrm{T}}\mathbf{X}\right)^{-1}$$

Meaningfulness of the regression terms

Regression terms meaningfulness significance test

- A test of significance can be done to understand if a **linear relation among predictors and response variable exists**:
  - hypothesis test

$$H_0\colon \beta_1 = \beta_2 = \cdots = \beta_V = 0$$
$$H_1\colon \beta_v \neq 0 \text{ for at least one } v$$

  - the rejection of H_0 implies that at least one of the **regressors contributes significantly to the model**
- The test procedure involves an analysis of variance partitioning the total sum of squares into a sum of squares due to the regression model and a sum of squares due to residual (or error):

$$SS_{\mathrm{T}} = SS_{\mathrm{R}} + SS_{\mathrm{E}}$$

  - where:
    - $SS_{\mathrm{R}} = \widehat{\boldsymbol{\beta}}^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}\mathbf{y} - \frac{\left(\sum_{n=1}^{N} y_n\right)^2}{N}$
    - $SS_{\mathrm{E}} = \mathbf{y}^{\mathrm{T}}\mathbf{y} - \widehat{\boldsymbol{\beta}}^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}\mathbf{y}$
    - $SS_{\mathrm{T}} = \mathbf{y}^{\mathrm{T}}\mathbf{y} - \frac{\left(\sum_{n=1}^{N} y_n\right)^2}{N}$
- If the null hypothesis is true, then:
  - SS_{R} is distributed like a χ^2_V distribution
  - SS_{E} is distributed like a χ^2_{N-V-1} distribution
  - SS_{E} and SS_{R} are independent
- The p-value approach can be used to reject the null hypothesis if:

$$F_o = \frac{MS_{\mathrm{R}}}{MS_{\mathrm{E}}} = \frac{SS_{\mathrm{R}}/V}{SS_{\mathrm{E}}/(N-V-1)} > F_{\alpha,V,N-V-1}$$

in this way one can verify if all the terms of the regression model meaningful

Test for a single regression coefficient

- It is always interesting to determine the **effectiveness of including/excluding one or more regressors** in the model
  - adding a variable to the regression model
    - always causes the sum of squares for regression to increase and the error sum of squares to decrease
      - is the increase in the regression sum of squares sufficient to warrant using the additional variable in the model?
    - if an unimportant variable is included into the model the mean square error can increase, thereby decreasing the usefulness of the model
- The hypotheses for testing the significance of any individual regression coefficient β_v are:

$$H_0: \beta_v = 0$$
$$H_1: \beta_v \neq 0$$

  - if $H_0: \beta_v = 0$ is not rejected, then this indicates that x_v can be deleted from the model
- The test statistic for this hypothesis is:

$$t_o = \frac{\hat{\beta}_v}{\sqrt{\hat{\sigma}^2 c_{vv}}} = \frac{\hat{\beta}_v}{SE(\hat{\beta}_v)}$$

  - where:
    - c_{vv} is the diagonal element of $\left(\mathbf{X}^{\mathrm{T}}\mathbf{X}\right)^{-1}$
    - $SE(\hat{\beta}_v)$ is the **standard error** of the regression coefficients
- The null hypothesis is rejected if:

$$|t_o| > t_{\alpha,N-V-1}$$

  - which is a **partial test** because the regression coefficient depends on all the other regressor variables in the model

Test for model regression coefficients

- The contribution to the regression sum of squares for a particular variable x_v given that all the other variables are included in the model may be tested, as well
- The **general regression significance test** (a.k.a. **extra sum of squares method**) is a procedure that goes through the following steps:
  - we would like to determine if **a subset of** L (with $L < V$) **regressors significantly contribute to the model**
  - let define $\boldsymbol{\beta} = \begin{bmatrix} \boldsymbol{\beta}_a \\ \boldsymbol{\beta}_b \end{bmatrix}$ where $\boldsymbol{\beta}_a$ is the set of L regression coefficients associated to the variables whose significance is to be assessed
  - hypothesis testing:

$$H_0: \boldsymbol{\beta}_a = \mathbf{0}$$
$$H_1: \boldsymbol{\beta}_a \neq \mathbf{0}$$

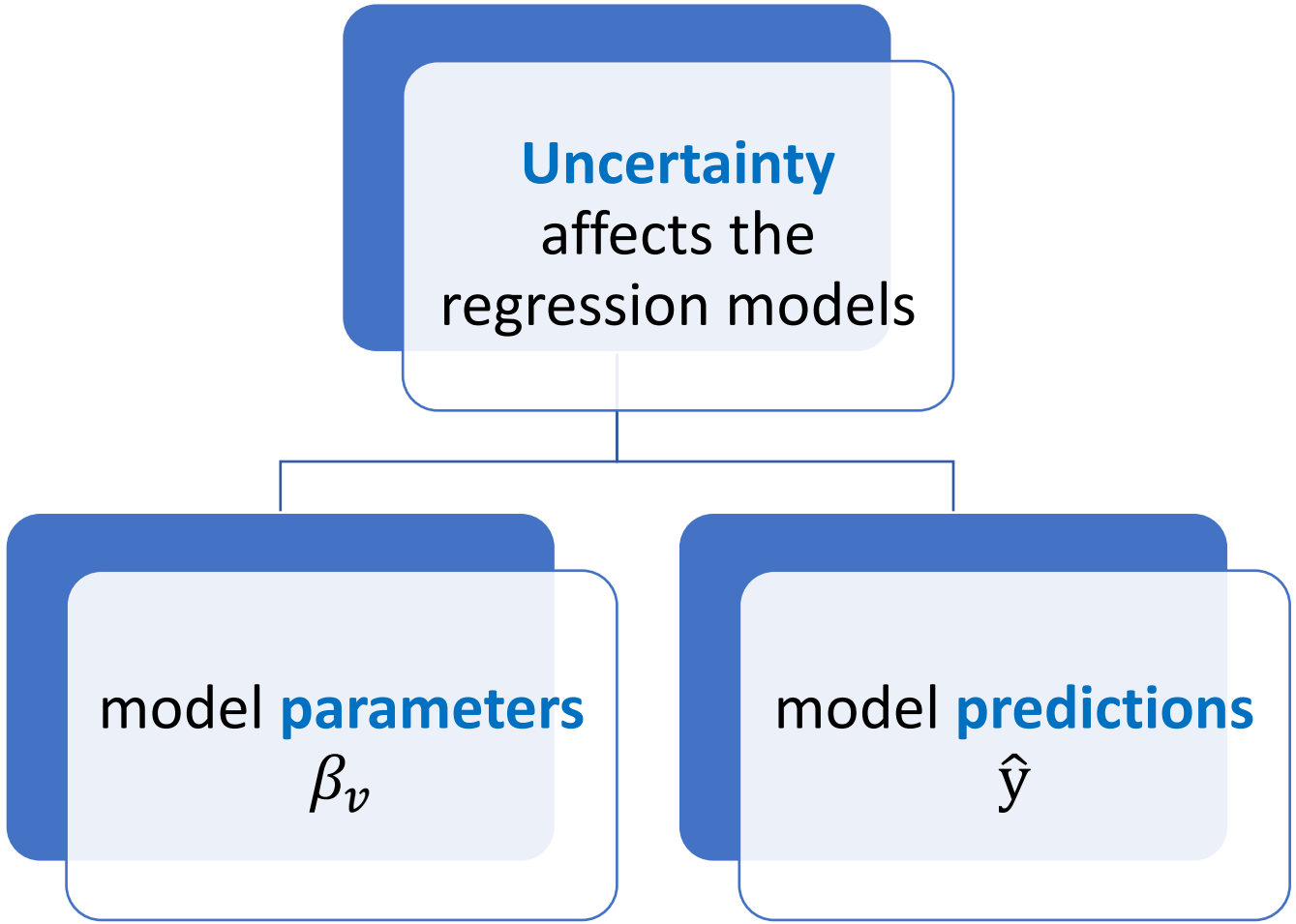
  - the test statistics defines the **partial F test**:

$$F_o = \frac{[SS_{\mathrm{R}}(\boldsymbol{\beta}) - SS_{\mathrm{R}}(\boldsymbol{\beta}_b)]/L}{MS_{\mathrm{E}}} > F_{\alpha,L,N-V-1}$$

    - where: $SS_{\mathrm{R}}(\boldsymbol{\beta}_b) = \hat{\boldsymbol{\beta}}_b^{\mathrm{T}} \mathbf{X}_b^{\mathrm{T}} \mathbf{y}$
    - **verifies if at least one of the L variables significantly contributes to the model**
    - this test on a single variable is the same as the t-test in the previous slide

Model uncertainty

Regression model uncertainty



Regression coefficients uncertainty

- Being $\widehat{\boldsymbol{\beta}}$ a linear combination of observations, they are normally distributed with mean vector $\boldsymbol{\beta}$ and covariance matrix $\sigma^2\left(\mathbf{X}^{\mathrm{T}}\mathbf{X}\right)^{-1}$, each of the statistics

$$\frac{\hat{\beta}_v - \beta_v}{\sqrt{\sigma^2 c_{vv}}}$$

  - is distributed as t with $(N - V)$ degrees of freedom
- The $100(1-a)\%$ percent **confidence interval for the regression coefficient** β_v is:

$$\hat{\beta}_v - t_{\alpha,N-V-1}\sqrt{\sigma^2 c_{vv}} \leq \beta_v \leq \hat{\beta}_v + t_{\alpha,N-V-1}\sqrt{\sigma^2 c_{vv}}$$

  - which is equivalent to:

$$\hat{\beta}_v - t_{\alpha,N-V-1}SE\left(\hat{\beta}_v\right) \leq \beta_v \leq \hat{\beta}_v + t_{\alpha,N-V-1}SE(\hat{\beta}_v)$$

Mean response uncertainty

- The $100(1-a)\%$ percent confidence interval for the mean response at the point $\mathbf{x}_n$ is:

$$\hat{y}(\mathbf{x}_n) - t_{\alpha,N-V}\sqrt{\hat{\sigma}^2\left[1+\mathbf{x}_n^{\mathrm{T}}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{x}_n\right]} \leq \boldsymbol{\mu}_y(\mathbf{x}_n)$$
$$\leq \hat{y}(\mathbf{x}_n) + t_{\alpha,N-V}\sqrt{\hat{\sigma}^2\left[1+\mathbf{x}_n^{\mathrm{T}}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{x}_n\right]}$$

  - where $\boldsymbol{\mu}_y(\mathbf{x}_n)$ is the mean response at $\mathbf{x}_n$
- A model that fits well in the region of the original data may no longer fit well outside of that region
  - for example, in a **confirmation experiment**, we are usually testing the model developed from the original experiment to determine if our interpretation was correct
  - often, we will do this by using the model to predict the response at some point of interest in the design space and then comparing the predicted response with an actual observation obtained by conducting another trial at that point
    - a useful measure of confirmation is to see if **the new observation falls inside the prediction interval on the response at that point**

Calibration data appropriateness

Data appropriateness for regression models

- Data may be inappropriate for building a regression model:
  - some data may be too “different” from the others to be included into the model
    - some data may derive from “wrong” measurements
    - outliers should be detected (and removed?)
  - some data may have too large an influence on the model
    - excessive leverage
- The study of the regression **model residuals** and **observation leverage** identifies abnormal observations

Scaled residuals

- **Scaled residuals** often convey more information than do the ordinary residuals
- For example, **standardized residual**:

$$d_n = \frac{e_n}{\hat{\sigma}}$$

  - have mean zero and approximately unit variance
  - are useful in looking for **outliers**
    - any observation with a standardized residual outside of the **interval [-3, 3]** is potentially unusual
- Residuals may have standard deviations that differ greatly. Scaling takes this into account
- The vector of fitted values corresponding to the observed values y_n is:

$$\hat{\mathbf{y}} = \mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}\left(\mathbf{X}^{\mathrm{T}}\mathbf{X}\right)^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{y} = \mathbf{H}\mathbf{y}$$

  - where $\mathbf{H} = \mathbf{X}\left(\mathbf{X}^{\mathrm{T}}\mathbf{X}\right)^{-1}\mathbf{X}^{\mathrm{T}}$ is the so-called **"hat" matrix** that maps the vector of observed values into a vector of fitted values

Studentized residuals

- The covariance matrix of the residuals (in a matrix notation) is

$$\text{cov}(\mathbf{e}) = \text{cov}(\boldsymbol{y} - \hat{\mathbf{y}}) = \sigma^2(\mathbf{I} - \mathbf{H})$$

  - generally, $(\mathbf{I} - \mathbf{H})$ is not diagonal, so the residuals have different variances and they are correlated
- The variance of the n-th residual is:

$$\text{V}(e_n) = \sigma^2(1 - h_{nn})$$

  - where h_{nn} is the n-th diagonal element of $\mathbf{H}$
    - because $h_{nn} \in [0, 1]$, using the residual mean square MS_E to estimate the variance of the residuals actually overestimates $\text{V}(e_n)$
    - because h_{nn} is a measure of the location of the n-th point in x-space, the variance of e_n depends on where the point $\mathbf{x}_n$ lies
      - residuals near the center of the x-space have larger variance than residuals at remote locations
      - violations of model assumptions are more likely at remote points
      - these violations may be hard to detect from inspection of residuals because their residuals will usually be smaller
- To take this inequality of variance into account the **studentized residuals** are suggested:

$$r_n = \frac{e_n}{\sqrt{\hat{\sigma}^2(1-h_{nn})}} \quad \text{with} \quad \hat{\sigma}^2 = MS_E$$

  - have constant unit variance regardless of the location of $\mathbf{x}_n$ when the form of the model is correct
  - *in many situations the variance of the residuals stabilizes, particularly for large data sets*
    - in these cases, there may be little difference between the standardized and studentized residuals
  - standardized and studentized residuals often convey equivalent information
  - examination of the studentized residuals is generally recommended because **any point with large residual and h_{nn} is highly influential on the least-squares fit**

Influence diagnostics

- **Point leverage**:
  - the disposition of points in x space is important in determining model properties
    - **remote observations potentially have disproportionate leverage** on the parameter estimates, predicted values, and the usual summary statistics
  - the **hat matrix $\mathbf{H}$ is very useful in identifying influential observations**
    - $\mathbf{H}$ determines the variances and covariance of the predicted value and the residuals
    - the elements of $\mathbf{H}$ may be interpreted as the amount of **leverage** exerted by y_n on $\hat{y}_n$
    - inspection of the elements of $\mathbf{H}$ can reveal points that are potentially influential by virtue of their location in the x space
  - as a rough guideline, if the n-th diagonal element of $\mathbf{H} > 2V/N$, observation n is a high-leverage point
- **Influence on regression coefficients**:
  - a measure of the squared distance between the least-squares estimate of the regression parameters $\hat{\boldsymbol{\beta}}$ based on all N observations and the estimate obtained by deleting the n–th point $\hat{\boldsymbol{\beta}}_{(n)}$:

$$D_n = \frac{r_n^2 h_{nn}}{V(1-h_{nn})}$$

    - a reasonable cutoff for D_n is unity ($\boldsymbol{D_n > 1}$ **is influential**)
    - note that D_n is the product of the square of the n-th studentized residual and the ratio of the distance from the vector $\mathbf{x}_n$ to the centroid of the remaining data
      - one component reflects **how well the model fits** the n-th observation
      - one component measures **how far that point is from the rest** of the data

Linear model structure appropriateness

Lack of fit testing

- **Repeated measurements** of the same point (i.e., replicates of the same experiment) allow the partitioning of the residual sum of squares SS_{E} into two components
  - the sum of squares due to **pure error** SS_{PE}
  - the sum of squares due to **lack of fit** SS_{LOF}
- Suppose that J_m observations are available for the *m*-th level of the regressors $\mathbf{x}_m$; the *mj*-th residual is:

$$y_{mj} - \hat{y}_m = \left(y_{mj} - \bar{y}_m\right) + \left(\bar{y}_m - \hat{y}_m\right)$$

  - where $\bar{y}_m$ is the average of J_m observations associated to the regressors $\mathbf{x}_m$
- Squaring both sides of the last equation and summing over *M* and J_m:

$$SS_{\mathrm{E}} = \sum_{m=1}^{M}\sum_{j=1}^{J_m}\left(y_{mj} - \bar{y}_m\right)^2 + \sum_{m=1}^{M} J_m\left(\bar{y}_m - \hat{y}_m\right)^2$$

  - **pure error sum of squares**: $SS_{\mathrm{PE}} = \sum_{m=1}^{M}\sum_{j=1}^{J_m}\left(y_{mj} - \bar{y}_m\right)^2$
    - is obtained by computing the corrected sum of squares of the repeat observations at each level and then pooling over the *M* levels
    - if the assumption of constant variance is satisfied, this is a model-independent measure of pure error
  - **sum of squares for lack of fit**: $SS_{\mathrm{LOF}} = \sum_{m=1}^{M} J_m\left(\bar{y}_m - \hat{y}_m\right)^2$
    - is a weighted sum of squared deviations between the mean response at each level and the corresponding fitted value
    - if the fitted values are close to the corresponding average responses, then there is a strong indication that the regression function is linear
    - if the predicted value deviates greatly from the real one, then it is likely that the regression function is not linear
- The test statistic for lack of fit is:

$$F_o = \frac{SS_{\mathrm{LOF}}/(M-V)}{SS_{\mathrm{PE}}/(N-M)} > F_{\alpha,M-V,N-M}$$

  - **if we conclude that the regression function is not linear, then the tentative model must be abandoned** and it is suggested to find a more appropriate equation
  - alternatively, there is no strong evidence of lack of fit

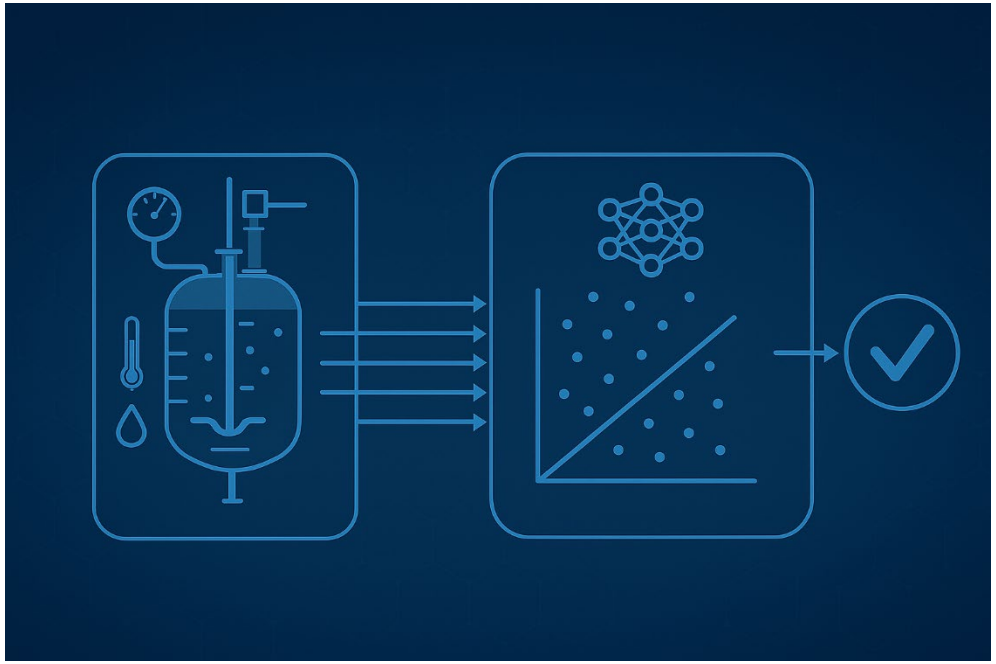
in this way one can verify if a linear model is valid

Homework

- Train with linear regression:
  - this will be one of the most important modelling method for the second part of the course on **Design of Experiments**
  - we will practice a lot with it!
- Consider Matlab® commands:
  - `regress`
  - `fitlm`
  - `fitrlinear`

Take-home message

- **Regression models** are used to make estimations/prediction of a response variable (e.g., the product quality), which is difficult to measure and can be obtained only at a low frequency, from some predictors (e.g., process variables) which are easy to measure and can be easily collected in real time, very frequently and at a low cost:
  - the **model adequacy** can be assessed through the determination coefficients
  - the **regressors** which are not important can e excluded thanks to appropriate hypothesis testing
  - the model **uncertainty** can be characterized both in terms of:
    - estimated/predicted response uncertainty
    - regression parameters uncertainty
  - the model **structure** can be verified through hypothesis testing



… per sempre a fianco a me!