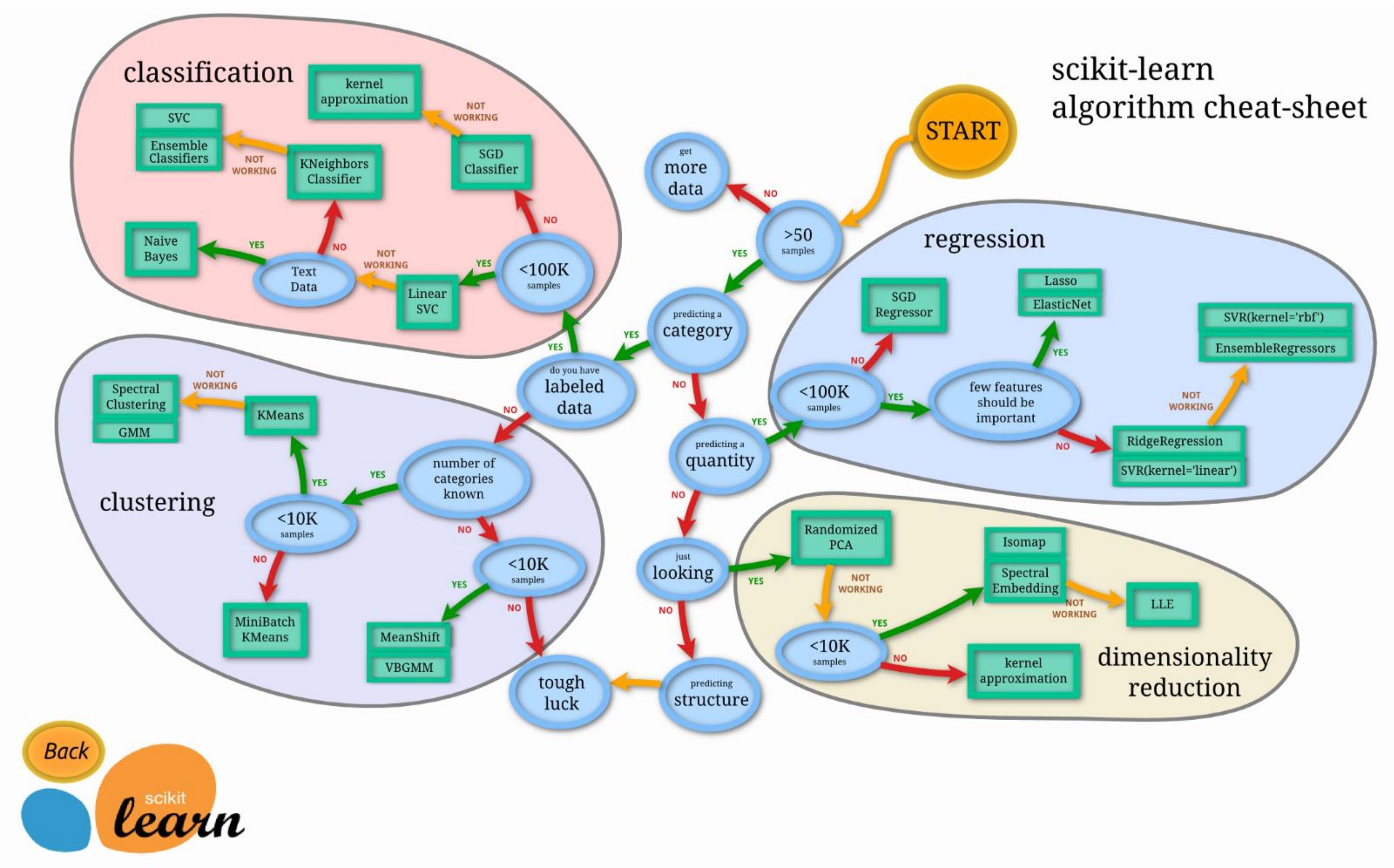
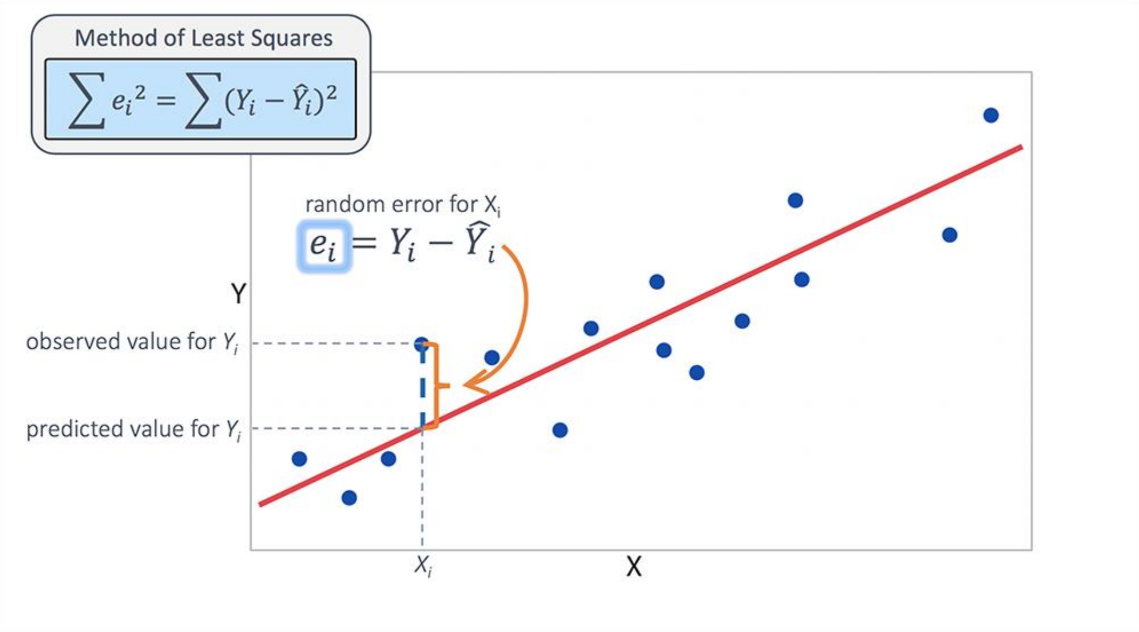
Lineær og Logistisk regression

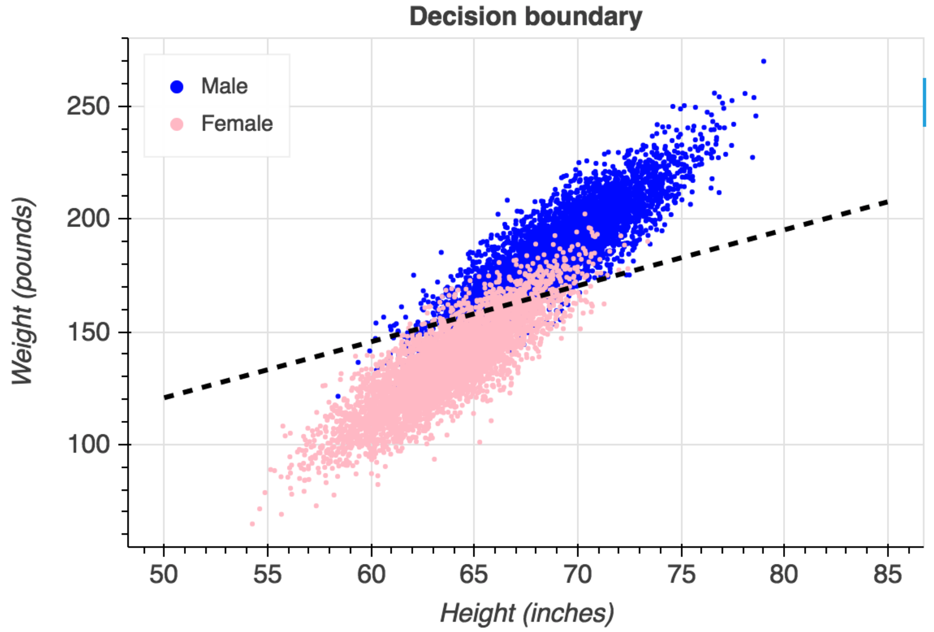
Introduktion til regression og klassifikation

Hvor er vi ? Regression og klassifikation..



Regression vs. klassifikation





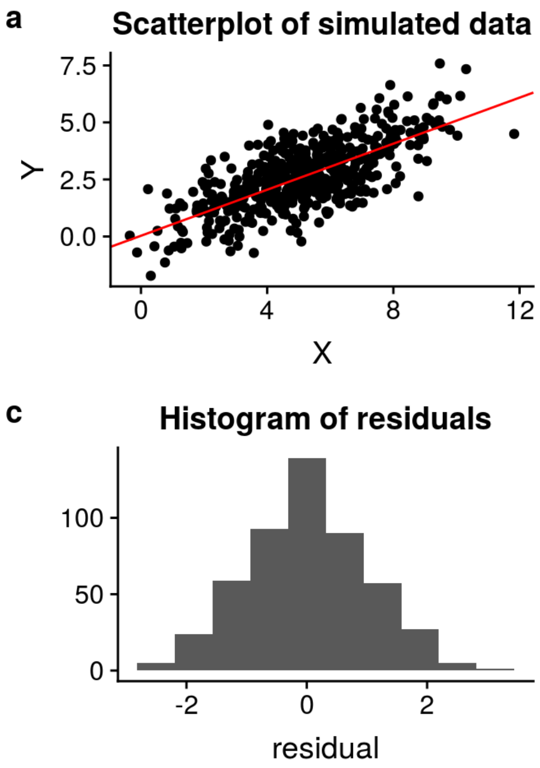
$$\boxed{Funktion\ y : R^D \rightarrow R}$$

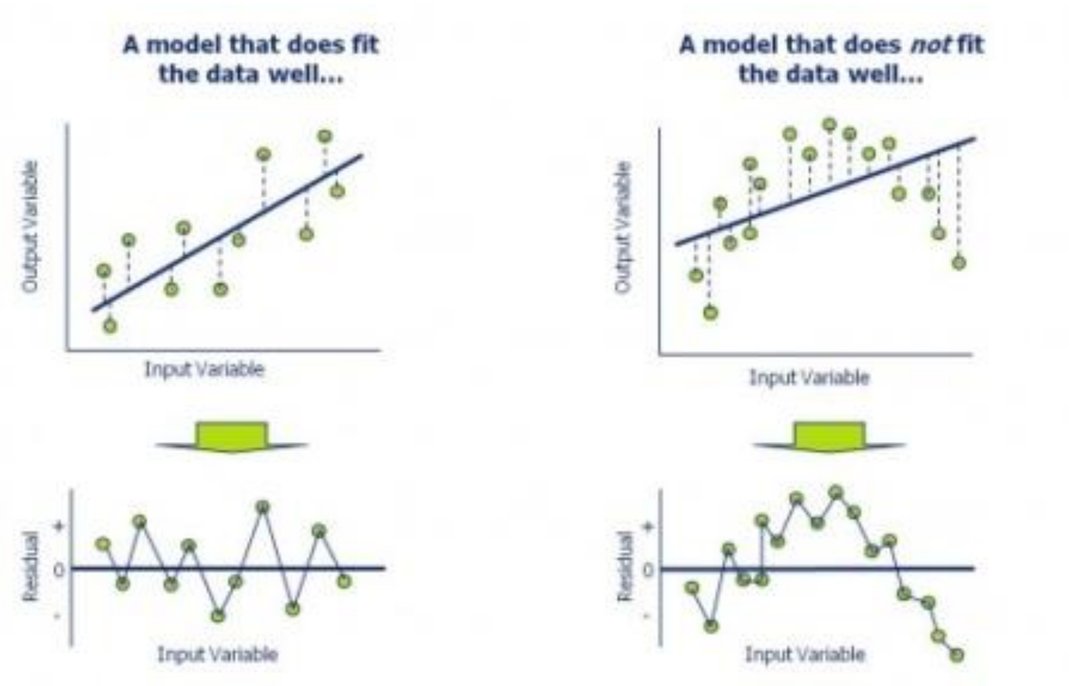
R = de reelle tal (kontinuert), D = dimension

$$\boxed{Funktion\ y : R^D \rightarrow C}$$

R = de reelle tal (kontinuert), D = dimension
C = diskrete tal {1, 2, …, antal klasser/kategorier}

Residual analyse





Lineær regression – til regression

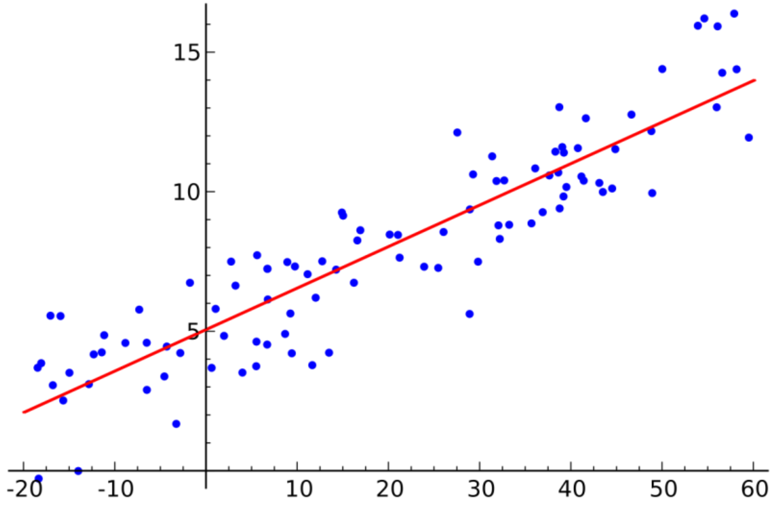
Equation 4-1. Linear Regression model prediction

$$\boxed{\hat{y} = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \cdots + \theta_n x_n}$$

Equation 4-3. MSE cost function for a Linear Regression model

$$\boxed{\mathrm{MSE}(\mathbf{X}, h_{\boldsymbol{\theta}}) = \frac{1}{m} \sum_{i=1}^{m} \left(\boldsymbol{\theta}^T \mathbf{x}^{(i)} - y^{(i)}\right)^2}$$

- $\hat{y}$ is the predicted value.
- n is the number of features.
- x_i is the i[th] feature value.
- θ_j is the j[th] model parameter (including the bias term θ_0 and the feature weights $\theta_1, \theta_2, \cdots, \theta_n$).



Vectorized form.. (bare lettere at skrive/læse..)

Equation 4-2. Linear Regression model prediction (vectorized form)

$$\hat{y} = h_{\boldsymbol{\theta}}(\mathbf{x}) = \boldsymbol{\theta} \cdot \mathbf{x}$$

- $\boldsymbol{\theta}$ is the model's *parameter vector*, containing the bias term θ_0 and the feature weights θ_1 to θ_n.
- $\mathbf{x}$ is the instance's *feature vector*, containing x_0 to x_n, with x_0 always equal to 1.
- $\boldsymbol{\theta} \cdot \mathbf{x}$ is the dot product of the vectors $\boldsymbol{\theta}$ and $\mathbf{x}$, which is of course equal to $\theta_0 x_0 + \theta_1 x_1 + \theta_2 x_2 + \cdots + \theta_n x_n$.
- $h_{\boldsymbol{\theta}}$ is the hypothesis function, using the model parameters $\boldsymbol{\theta}$.

Løsning – Closed-form solution

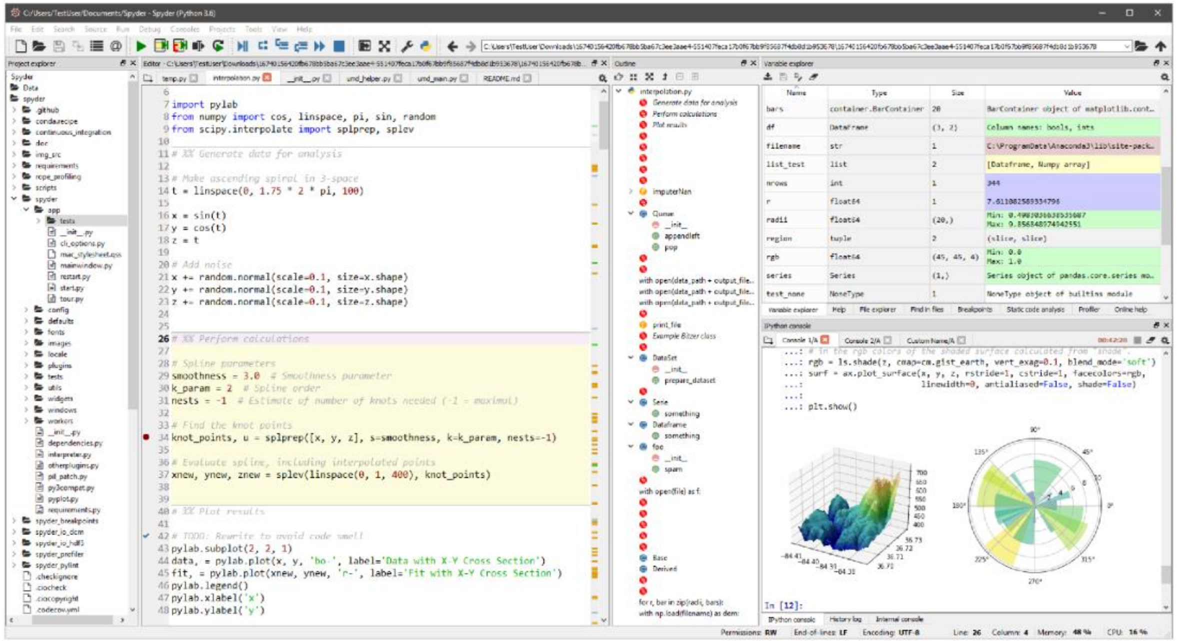
Equation 4-4. Normal Equation

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

- $\hat{\boldsymbol{\theta}}$ is the value of $\boldsymbol{\theta}$ that minimizes the cost function.
- $\mathbf{y}$ is the vector of target values containing $y^{(1)}$ to $y^{(m)}$.

Hands-on Python - Spyder

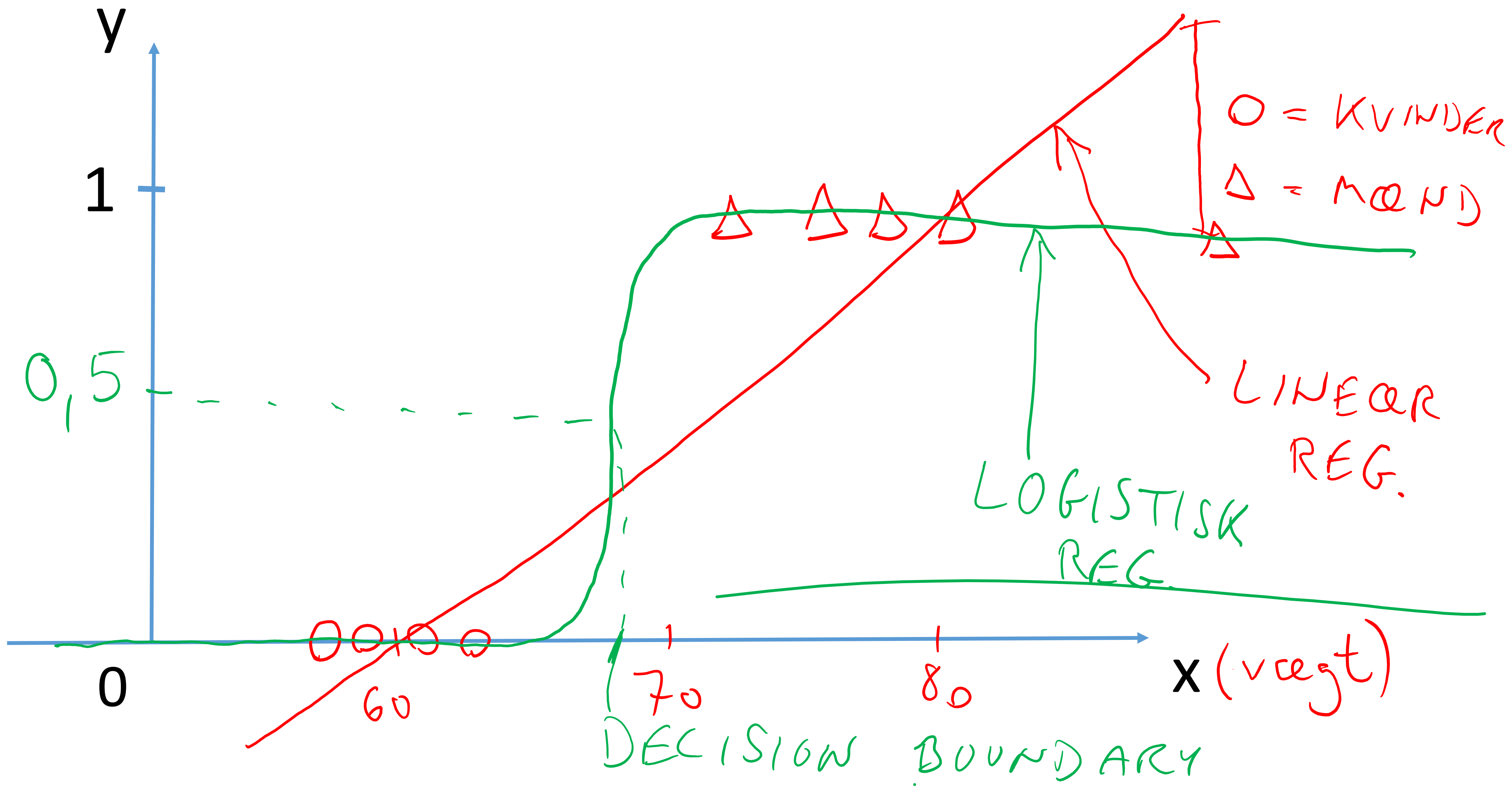
- https://www.spyder-ide.org/
- IDE til data science
- Minder meget om Matlab
- Del af Anaconda distribution



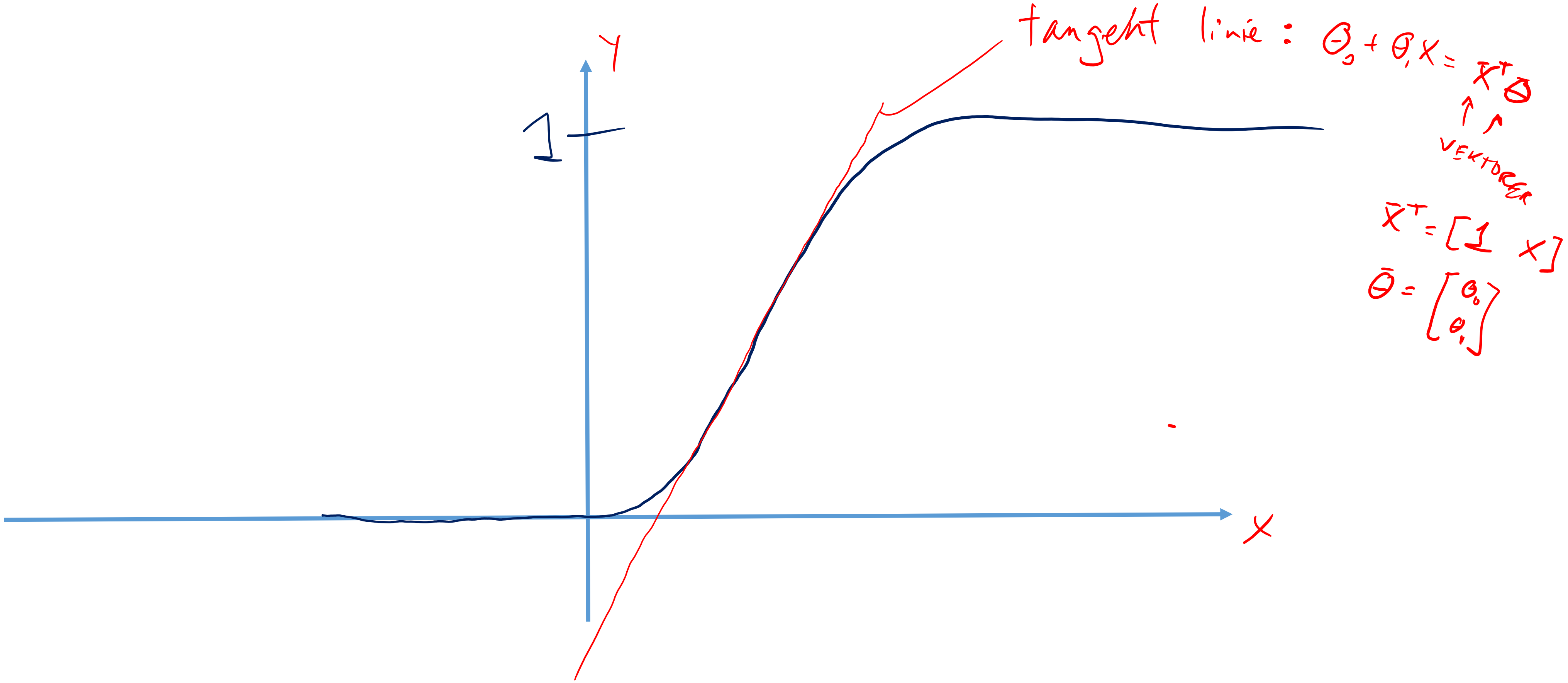
Hands-on Python

- Introduktion – om linear model
  - https://scikit-learn.org/dev/modules/linear_model.html
- Eksempel
  - https://scikit-learn.org/dev/auto_examples/linear_model/plot_ols.html#sphx-glr-auto-examples-linear-model-plot-ols-py

Lineær og logistisk regression til klassifikation



Logistisk regression funktionen



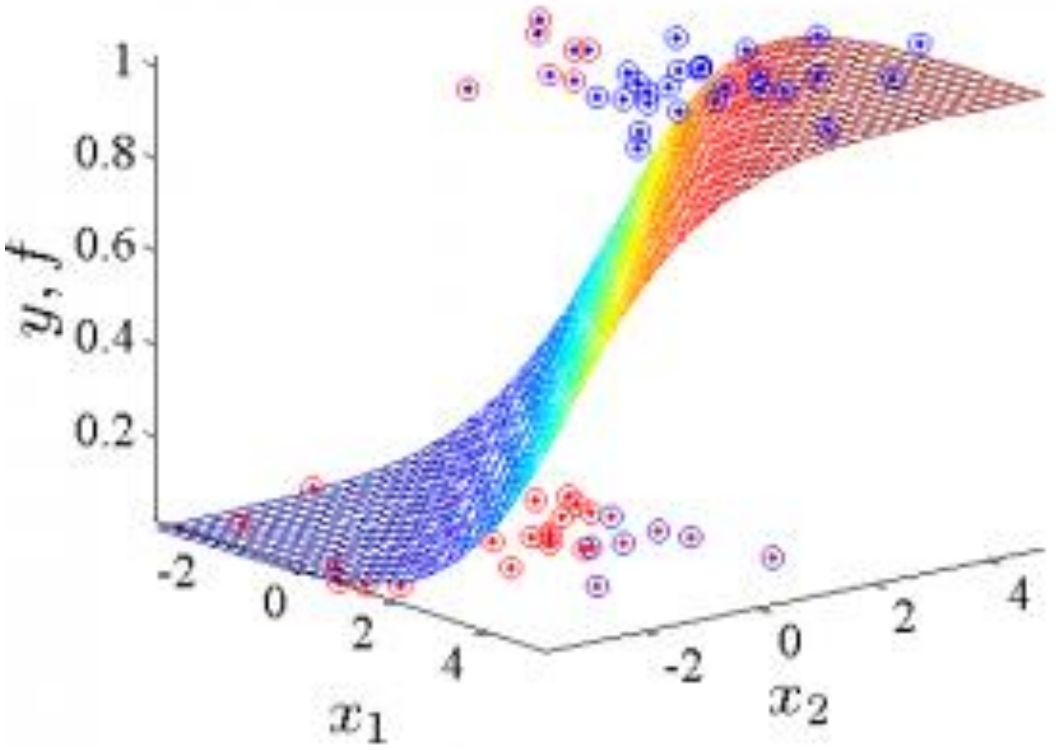
Equation 4-13. Logistic Regression model estimated probability (vectorized form)

$$\hat{p} = h_{\boldsymbol{\theta}}(\mathbf{x}) = \sigma\left(\mathbf{x}^T \boldsymbol{\theta}\right)$$

Equation 4-14. Logistic function

$$\sigma(t) = \frac{1}{1 + \exp(-t)}$$

Logistisk regression – 2D



Training / learning for logistisk regression

Equation 4-13. Logistic Regression model estimated probability (vectorized form)

$$\hat{p} = h_{\boldsymbol{\theta}}(\mathbf{x}) = \sigma\left(\mathbf{x}^T \boldsymbol{\theta}\right)$$

Equation 4-17. Logistic Regression cost function (log loss)

$$J(\boldsymbol{\theta}) = -\frac{1}{m}\Sigma_{i=1}^{m}\left[y^{(i)}log\left(\hat{p}^{(i)}\right) + \left(1 - y^{(i)}\right)log\left(1 - \hat{p}^{(i)}\right)\right]$$

Kan ikke minimeres med closed-form solution

Iterative solution – Gradient descent

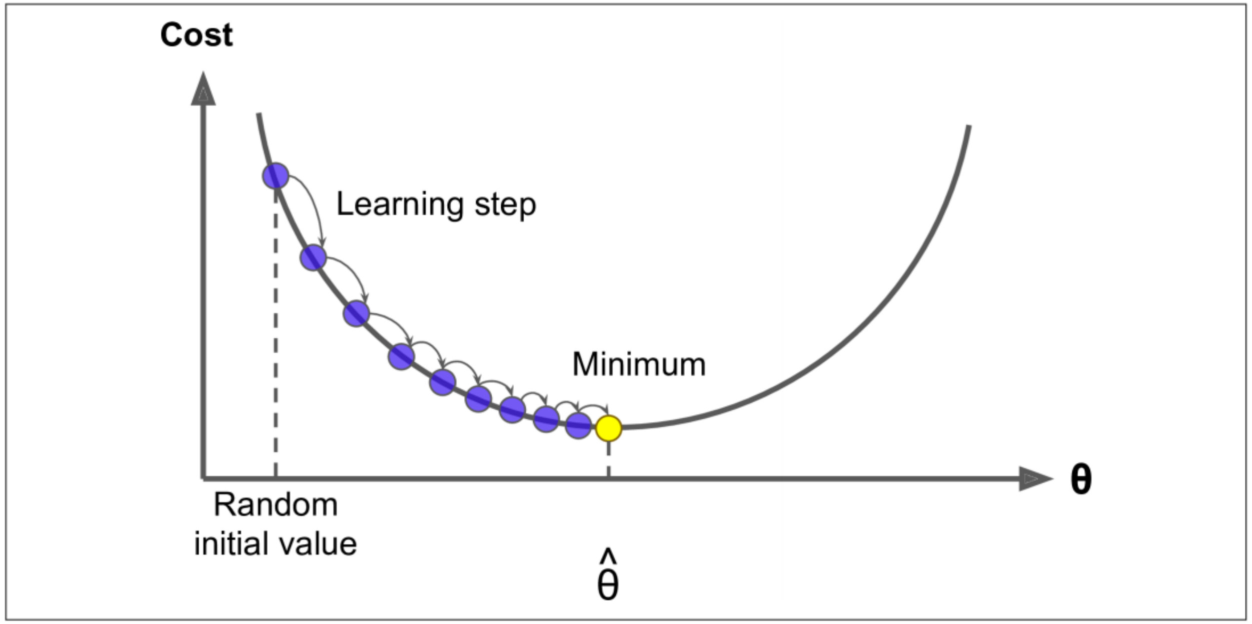


Figure 4-3. Gradient Descent

Equation 4-18. Logistic cost function partial derivatives

$$\frac{\partial}{\partial \theta_j} \mathrm{J}(\boldsymbol{\theta}) = \frac{1}{m} \sum_{i=1}^{m} \left(\sigma\left(\boldsymbol{\theta}^T \mathbf{x}^{(i)} \right) - y^{(i)} \right) x_j^{(i)}$$

Learning rate

Equation 4-7. Gradient Descent step

$$\boldsymbol{\theta}^{(\text{next step})} = \boldsymbol{\theta} - \eta \nabla_{\boldsymbol{\theta}} \mathrm{MSE}(\boldsymbol{\theta})$$

Gradient descent – learning rate

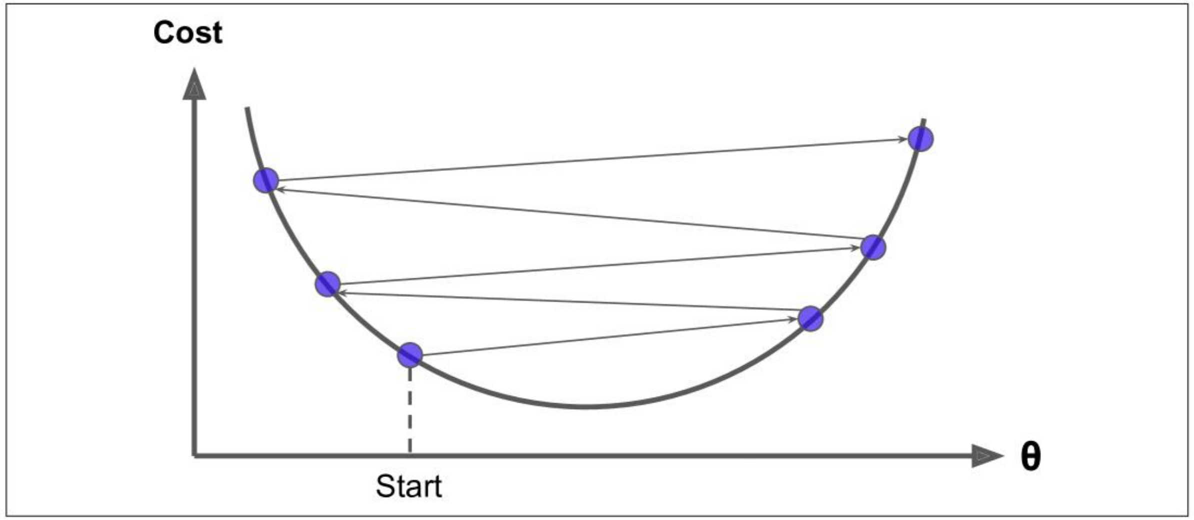


Figure 4-5. Learning rate too large

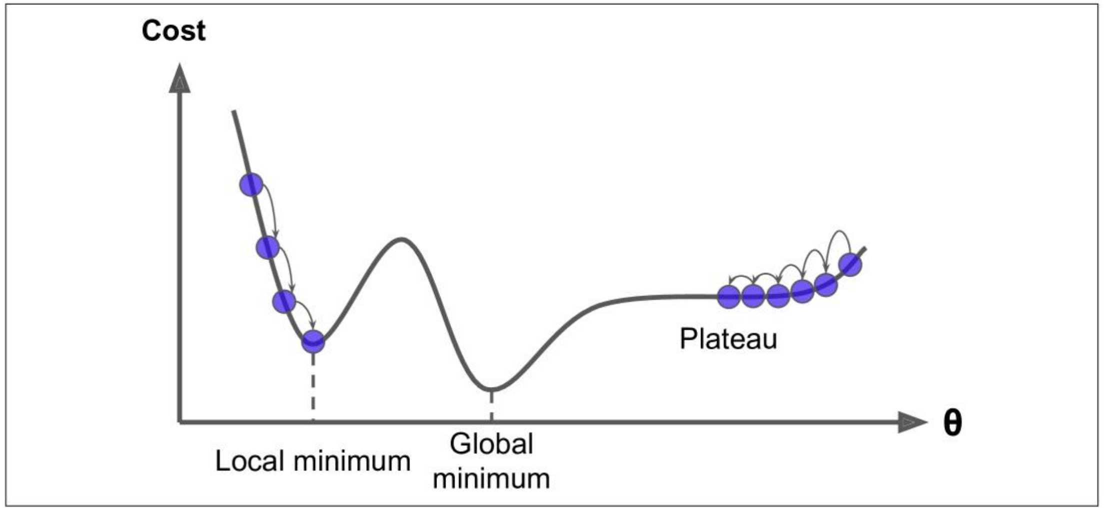
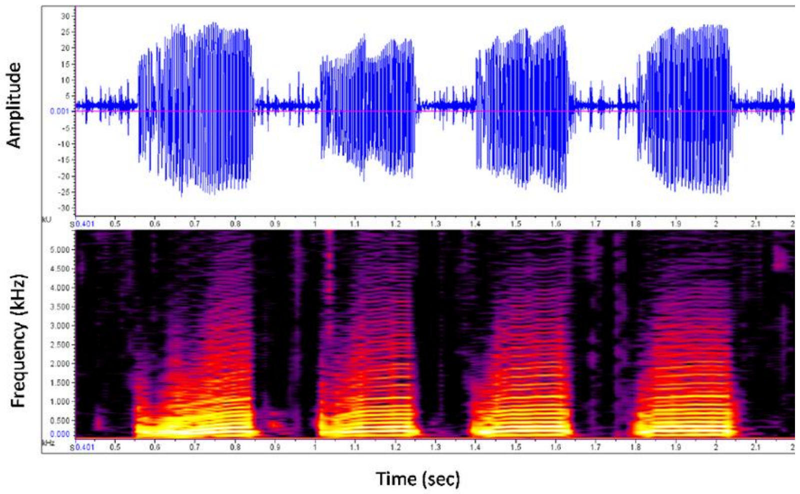
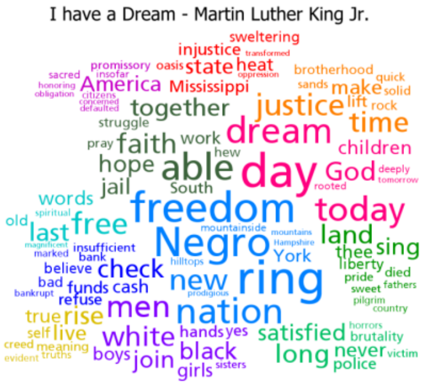
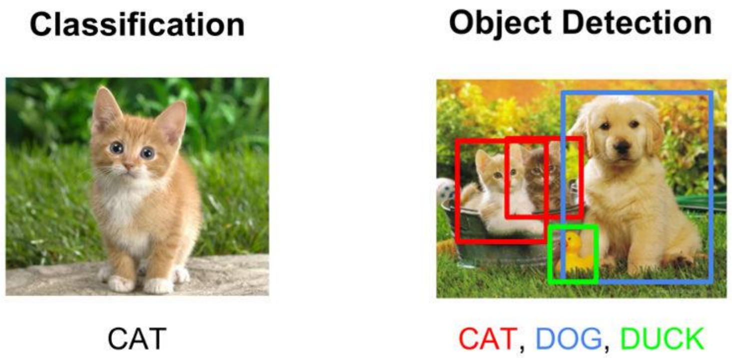


Figure 4-6. Gradient Descent pitfalls

Valg af eget projekt



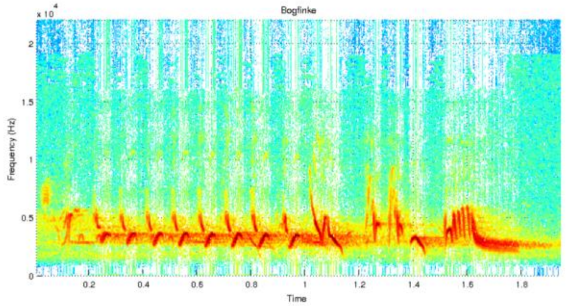




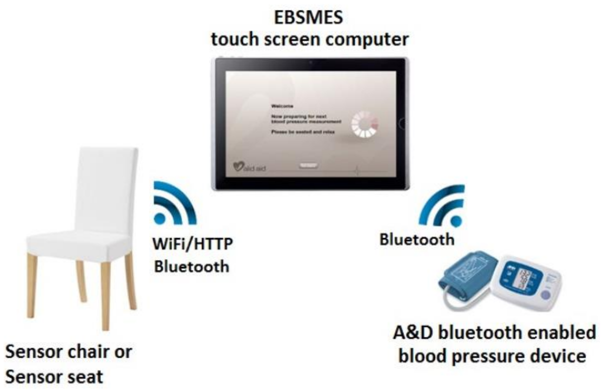
kaggle

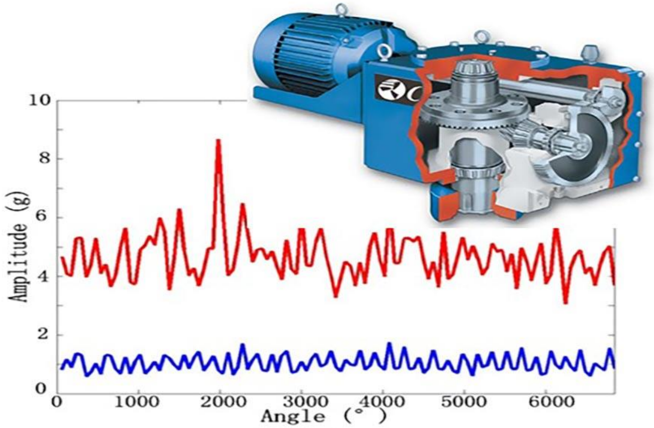
UCI Machine Learning Repository

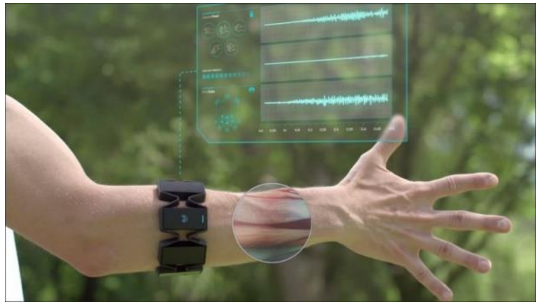
Eksempler - anvendelser



Figur 1.3: Spektrogram af bogfinken







Repræsentation af billeder/lyd/tekst/…