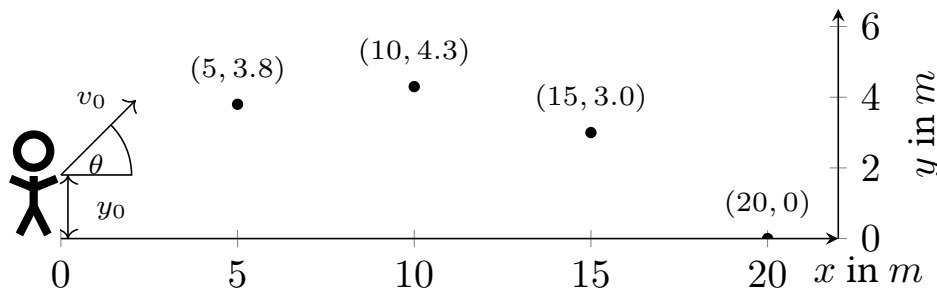


Deadline: Friday November 13th, 10:00 Please upload a .pdf file via LearnWeb. (e.g. exported Jupyter notebook)

1 Multidimensional Linear Regression

(10 points)



Tom throws a ball and the trajectory is recorded with a camera. From classical mechanics we know that the trajectory is of the form

$$y(x) = y_0 + \tan(\theta)x - \frac{g}{2v_0^2 \cos^2(\theta)}x^2 = \beta_0 + \beta_1x + \beta_2x^2$$

where y_0 is the initial height (in m), v_0 is the initial velocity (in $\frac{m}{s}$) and θ is the angle (in radians).

A [7p] Fit a linear regression model $\hat{y}(x) = \beta_0 + \beta_1x + \beta_2x^2 = \begin{pmatrix} 1 & x & x^2 \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix}$.

B [3p] Given that the gravitational constant is $g = 9.81 \frac{m}{s^2}$, estimate y_0 , v_0 and θ . What is the relative error, given that the true values are $y_0 = 1.8m$, $\theta = 30^\circ$ and $v_0 = 14 \frac{m}{s}$?

2 Gradient Descent – Programming

(10 points)

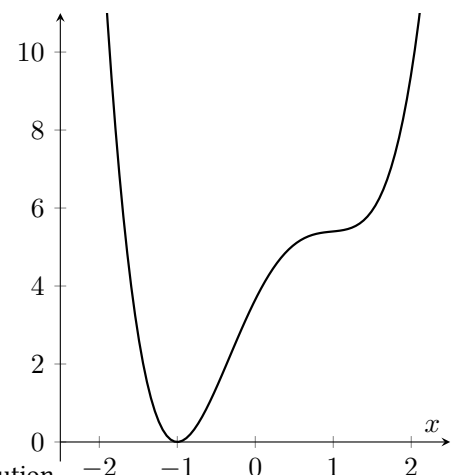
In this exercise we want to study the behaviour of gradient descent on the test function $f(x) = x^4 - 1.3x^3 - 1.95x^2 + 4x + 3.65$. This function is non-negative and has precisely one local (and also global) minimum at $x^* = -1$, $f(x^*) = 0$.

A [3p] Perform gradient descent with starting point $x_0 = 2$ and step length $\alpha = 0.1$. How many iterations are needed until the function value drops below 10^{-6} ?

B [3p] Try again with the same step length, but from the starting point $x_0 = 1.5$. Why does it take more iterations to achieve the target accuracy, although the starting point is closer to the minimum?

C [1p] What happens when the starting point $x_0 = -0.5$ with step-length $\alpha = 0.15$ is chosen?

D [3p] Repeat 1-3, but using backtracking line search, using $\alpha = 0.1$ and $\beta = 0.5$. How many iterations are needed in each case? What happens if α is chosen too small or too large (e.g. $\alpha = 0.95$)?



[4] Bonus Problems

[2] Prove that the normal equation $X^T X \beta = X^T y$ always has at least 1 solution.

[2] Prove that the solution is unique if and only if the features, i.e. the columns of the data matrix $X \in \mathbb{R}^{NM}$ (with $N \gg M$), are linearly independent.

Hint: First show part B by proving $\text{Ker}(X^T X) = \text{Ker}(X)$. Then prove A by showing $\text{Im}(X^T X) = \text{Im}(X^T)$ through application of the fundamental theorem $\text{Im}(X)^{\perp} = \text{Ker}(X^T)$.