**Task 58: Script Chapter 3 (Linear Regression) Summary**

Linear regression: Linear approach to model the relationship between **dependent variable** (target or label) and one or more **independent variables** (regressors or explanatory variables).

* Multivariate data (tuples of two variables)

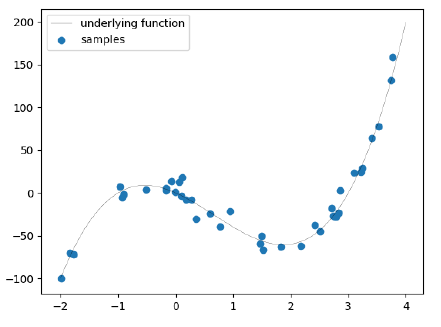
Dependent variable: Vector/ scalar: Multiple samples lead to multiple values of the dependent variable t1, t2, …, tn.

Independent variable: Vector/ scalar ~x1, ~x2, ..., ~xn: Each of these vectors has the same dimension, e.g. ~xi = (xi,1, xi,2, …, xi,j ).

**The underlying model and noise**

For the multivariate data there is some underlying function f(~x) (ti = f(~xi)) in the real world exactly governing the relationship, this is the function we are trying to approximate.

* function governing the relationship between ~xi and ti
* sampled points will usually not lie exactly on the function plot -> deviation from the underlying function due to noise

Idea: ti = f(~xi) + epsilon, where epsilon is the noise which is modelled as a normal distribution that is centred around 0 with variance sigma2

ti = Nmü=f(~xi),sigma2 (ki)

**Vanilla linear regression model**

yi = w0 \* 1 + w1 \* xi,1 + w2 \* xi,2 + … + wj \* xi,j = ~w \* ~xT

~w = weights (parameters of our model)

yi = approximated value for the dependent variable ti

The linear regression model assumes the following:

Homoscedasticity: Each sample has the same variance in its noise like the others, regardless of the values contained in this sample.

Independence of errors: The noise (error) of the samples is not correlated with each other.

* no non-linearity in the model, we can only model linear relationships between the independent and the dependent values -> **linear basis function models** to the rescue!

**Linear Basis Function Model**

Instead of linearly combining the "vanilla" independent variables, we linearly combine a set of simple non-linear functions (**basis functions**) of the independent variables. This set of simple functions is called the **basis set**. Since this is a linear basis function model in terms of the combination of coefficients, the coefficients themselves may still be nonlinear.

yi = w0 \* phi0(1) + w1 \*phi1(xi,1) + … + wj \* phij(xi,j) = ~w \* ~phi(~x)T

**Finding the right weights**

The weights are normal distributed in the form Nmü=f(~xi), sigma2 (ki) which is somehow equivalent to finding the underlying function. The underlying function is modelled as a linear basis function model of the form g(~x) = ~wT \* ~phi(~xi). Since we have multiple parameters, we have to take the derivative of the log likelihood (**score function**) in respect to the gradient.