Normal GLM

We have a data set containing a dependent variable y and a covariate x. We suspect that there is a quadratic dependence of y on x, so we start with the standard linear model

$$Y_i = \beta_0 + \beta_1 x_i + \beta_2 x^2 + U_i,$$

where $U_i \sim N(0, \sigma^2)$. In the file `Normal_GLM.Rdata` you can find $n = 50$ realizations of x and y.

(a) Fit the model and test whether the quadratic term is significant at 5% significance. Furthermore, in the quadratic model, make a plot of the residuals and discuss whether the data is homoscedastic.

There are theoretical reasons to suspect that the model is heteroscedastic. Therefore, we suggest to fit the model where

$$Y_i \sim N(\mu_i, \sigma_i^2),$$

with the natural parameters

$$\eta_i = (\eta_{i1}, \eta_{i2}) = (\frac{\mu_i}{\sigma_i^2}, \frac{1}{\sigma_i^2}).$$

We model the natural parameters with a linear model:

$$\eta_{i1} = \beta_{10} + \beta_{11} x_i + \beta_{12} x_i^2 \text{ and } \eta_{i2} = \beta_{20} + \beta_{21} x_i + \beta_{22} x_i^2.$$

In the script `Normal_GLM.r` you can find the three functions `lik`, `Dlik` and `D2lik`, that calculate the log-likelihood, the gradient and the second derivative matrix of the log-likelihood corresponding to this model respectively. Read through the functions to see how you should use them.

(b) Calculate the log-likelihood of the data using the function `lik`, for the parameters corresponding to the model you fitted in (a) (this is possible because model (a) is a submodel of the normal GLM: explain!). Think about how to define the matrices `X1` and `X2`. Also calculate the gradient and discuss your findings (the gradient has a special feature that you can explain!).

Now we wish to find the MLE under the normal GLM model. The log-likelihood function $l(\beta_1, \beta_2)$ is a concave function, and we know its gradient and second derivative. We can apply a variant of the Newton-Raphson optimization method: we start at an initial parameter β, and calculate $l(\beta), Dl(\beta)$ and $D^2 l(\beta)$. Now we apply a Taylor approximation:

$$l(\beta + h) = l(\beta) + Dl(\beta)h + \frac{1}{2} h^t D^2 l(\beta) h + o(\|h\|^2).$$

We can maximize the quadratic function in h and find

$$h_m = -(D^2 l(\beta))^{-1} Dl(\beta)^t.$$

Now calculate $l(\beta + h_m)$ and check if it is bigger than $l(\beta)$. If not, then divide h_m by 2 (so replace h_m by $h_m/2$) and check again if $l(\beta + h_m)$ is bigger than $l(\beta)$. Repeat this until you find the next value of the parameter with larger value for l, and hen repeat the whole procedure until you met a certain criterion. This could be a very small gradient or a very small increase of the function l.

(c) Implement the Newton-Raphson optimization to find the MLE of (β_1, β_2). Be careful: not all values for β_2 are allowed, since η_2 has to be positive! If you do not succeed, use the `R`-function `optim` to find the MLE. Note that in `optim`, you can also give the gradient function! Plot the data together with the estimated expectation according to the normal GLM model and according to the standard linear model. Also plot the residuals (using GLM) with plus and minus the estimated standard error at each value of x. You could also make this last plot for the standard linear model.

(d) Test whether the quadratic term in η_1 is significant, and test whether the quadratic term in η_2 is significant, both at the 5% level. Furthermore, test whether the normal GLM is a better model for this data than the standard normal model.

(e) We wish to see whether the normal GLM model can be used effectively for prediction. For this, we repeat the following experiment many times: simulate new y_i's using the model fitted in (c). Then fit the GLM model, and predict the value of y at $x = 0.75$. Also determine the standard deviation at $x = 0.75$. Do the same prediction, but now using the standard linear model. Compare both predictions with the (known!) true value, given by the model you simulate from. Determine which prediction is best. Also, think of a way to assess the estimation of the standard deviation at $x = 0.75$. If you feel up to it, you could repeat this for different vaues of x!