Solutions to Part A of Problem Sheet 3

Solution (3.1) We first compute the gradient and the Hessian of this function.

$$\nabla f(x_1, x_2) = \frac{1}{\sqrt{1+x^2+y^2}} \begin{pmatrix} x \\ y \end{pmatrix}, \quad (1)$$

$$\nabla^2 f(x_1, x_2) = \frac{1}{(1+x^2+y^2)^{3/2}} \begin{pmatrix} 1+y^2 & -xy \\ -xy & 1+x^2 \end{pmatrix}.$$

We have a stationary point at $(0,0)$ which is a minimizer, as the function can never fall below $f(0,0) = 1$. This means that the Hessian is positive definite at $(0,0)$. There are various ways of verifying that the Hessian is positive definite everywhere, and the function therefore convex. One is direct verification:

$$\boldsymbol{v}^\top \nabla^2 f(x,y) \boldsymbol{v} = v_1^2(1+y^2) - 2v_1v_2xy + v_2^2(1+x^2) = v_1^2 + v_2^2 + (v_1y_1 - v_2x_2)^2 > 0.$$

Newton's method starts with a point $(x_{(0)}, y_{(0)})$, and then for every $k \geq 0$, first solves the system of equations

$$\frac{1}{(1+x_{(k)}^2+y_{(k)}^2)^{3/2}} \begin{pmatrix} 1+y_{(k)}^2 & -x_{(k)}y^{(k)} \\ -x_{(k)}y_{(k)} & 1+x_{(k)}^2 \end{pmatrix} \begin{pmatrix} \Delta x \\ \Delta y \end{pmatrix} = \frac{1}{\sqrt{1+x_{(k)}^2+y_{(k)}^2}} \begin{pmatrix} x_{(k)} \\ y_{(k)} \end{pmatrix},$$

and then computes

$$(x_{(k+1)}, y_{(k+1)}) = (x_{(k)}, y_{(k)}) + (\Delta x, \Delta y).$$

Solution (3.2) We have the following four objective functions (two models on two data sets):

Table 1:

$$\begin{aligned} f_1(x_1, x_2) &= \ell(\beta_0+\beta_1+\beta_2-1) \\ &+ \ell(\beta_0+\beta_1-\beta_2+1) \\ &+ \ell(\beta_0-\beta_1+\beta_2+1) \\ &+ \ell(\beta_0-\beta_1-\beta_2-1), \end{aligned} \qquad \begin{aligned} f_2(x_1, x_2) &= \ell(\beta_0+\beta_1+\beta_2+\beta_3-1) \\ &+ \ell(\beta_0+\beta_1-\beta_2-\beta_3+1) \\ &+ \ell(\beta_0-\beta_1+\beta_2-\beta_3+1) \\ &+ \ell(\beta_0-\beta_1-\beta_2+\beta_3-1), \end{aligned}$$

Table 2:

$$\begin{aligned} f_3(x_1, x_2) &= \ell(\beta_0+\beta_1+\beta_2-1) \\ &+ \ell(\beta_0+\beta_1-\beta_2+1) \\ &+ \ell(\beta_0-\beta_1+\beta_2-1) \\ &+ \ell(\beta_0-\beta_1-\beta_2+1), \end{aligned} \qquad \begin{aligned} f_4(x_1, x_2) &= \ell(\beta_0+\beta_1+\beta_2+\beta_3-1) \\ &+ \ell(\beta_0+\beta_1-\beta_2+\beta_3+1) \\ &+ \ell(\beta_0-\beta_1+\beta_2+\beta_3-1) \\ &+ \ell(\beta_0-\beta_1-\beta_2+\beta_3+1), \end{aligned}$$

If we use the loss function $\ell(x) = \mathbf{1}\{x \neq 0\}$, then we need to choose the β_i in such a way as to make as many of the terms in the above formulas zero. To figure out whether it is possible to fit the data perfectly, we attempt to solve a system of linear equations for each of the cases above.

For f_1, we have zero loss (perfect fit) if:

$$\begin{aligned}\beta_0 + \beta_1 + \beta_2 - 1 &= 0\\ \beta_0 + \beta_1 - \beta_2 + 1 &= 0\\ \beta_0 - \beta_1 + \beta_2 + 1 &= 0\\ \beta_0 - \beta_1 - \beta_2 - 1 &= 0\end{aligned}$$

This system of equations has no solution. To see why, note that adding all equations together gives $\beta_0 = 0$, while adding the middle two gives $\beta_0 = -1$. Can we satisfy three of these equations? Yes, solving the first three equations gives $\beta_0 = -1, \beta_1 = 1, \beta_2 = 1$. It follows that the function

$$h_1(x_1, x_1) = -1 + x_1 + x_2$$

has loss 1 on the data of table 1 (it makes an error only on the last row).

For f_2, we have zero loss (perfect fit) if:

$$\begin{aligned}\beta_0 + \beta_1 + \beta_2 + \beta_3 - 1 &= 0\\ \beta_0 + \beta_1 - \beta_2 - \beta_3 + 1 &= 0\\ \beta_0 - \beta_1 + \beta_2 - \beta_3 + 1 &= 0\\ \beta_0 - \beta_1 - \beta_2 + \beta_3 - 1 &= 0\end{aligned}$$

We can solve this system of equations by row elimination, or by noting that $\beta_3 = 1$ and $\beta_0 = \beta_1 = \beta_2 = 0$ works. Thus

$$h_2(x_1, x_2) = x_1 x_2$$

describes the data perfectly! (As could have been guessed by looking at the table: the y value is the product of the x_1 and x_2 values).

For f_3, we have zero loss (perfect fit) if:

$$\begin{aligned}\beta_0 + \beta_1 + \beta_2 - 1 &= 0\\ \beta_0 + \beta_1 - \beta_2 + 1 &= 0\\ \beta_0 - \beta_1 + \beta_2 - 1 &= 0\\ \beta_0 - \beta_1 - \beta_2 + 1 &= 0\end{aligned}$$

Again, we see that $\beta_0 = \beta_1 = 0$ and $\beta_2 = 1$ is a solution, so that

$$h_1(x_1, x_2) = x_2$$

fits the data of table 2 perfectly. Finally, for f_4 we only need to observe that f_3 is a special case, with $\beta_3 = 0$, so that $h_2(x_1, x_2) = x_2$ also holds in this case.

To summarize: to fit the data from table one perfectly, we need a quadratic function, while a linear function is good for fitting the data from table 2.