fake title

Jesper Løve Hinrich
912933169

5. maj 2014

1 The derivative of prediction or Sensitivity

We wish to find the effect that a datapoint's class has on the predicted class for that datapoint.

$$\frac{\delta \hat{Y}_n}{\delta Y_n} \tag{1.1}$$

Our prediction is

$$\hat{Y}_n = p(y|\bar{x}, \bar{w}) \tag{1.2}$$

where $\bar{w}$ is subject to

$$\frac{\delta L}{\delta \bar{w}} = 0 \tag{1.3}$$

Which means that we have found a locally optimal solution.

We now assume that when we move y by a small amount δy then 1.3 still holds.

Essentially assuming some smoothness around the optimum.

Using this and the fact that 1.3 depends both directly and indirectly on y we see that

$$\frac{\delta}{\delta y}\frac{\delta L}{\delta w} = 0$$

$$\Downarrow$$

$$\frac{\delta^2 L}{\delta y \delta \bar{w}} + \frac{\delta^2 L}{\delta \bar{w} \delta \bar{w}^T}\frac{\delta \bar{w}}{\delta y} = 0$$

and from this we can isolate

$$\frac{\delta \bar{w}}{\delta y} = -\left[\frac{\delta^2 L}{\delta \bar{w} \delta \bar{w}^T}\right]^{-1} \frac{\delta^2 L}{\delta y \delta \bar{w}} \tag{1.4}$$

Rewriting 1.1 we get

$$\frac{\delta \hat{Y}_n}{\delta Y_n} = \frac{\delta p(y|\bar{x}, \bar{w})}{\delta Y_n} = \frac{\delta p(y|\bar{x}, \bar{w})}{\delta \bar{w}^T}\frac{\delta \bar{w}}{\delta y} \tag{1.5}$$

And inserting 1.4

$$\frac{\delta p(y|\bar{x},\bar{w})}{\delta \bar{w}^T}\frac{\delta \bar{w}}{\delta y} = -\frac{\delta p(y|\bar{x},\bar{w})}{\delta \bar{w}^T}\left[\frac{\delta^2 L}{\delta \bar{w}\delta \bar{w}^T}\right]^{-1}\frac{\delta^2 L}{\delta y \delta \bar{w}} \tag{1.6}$$