fake title

Jesper Løve Hinrich
912933169

6. 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 **??** still holds.

Essentially assuming some smoothness around the optimum.

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

$$\begin{gathered}
\tfrac{\delta}{\delta y}\tfrac{\delta L}{\delta w} = 0 \\
\Downarrow \\
\tfrac{\delta^2 L}{\delta y \delta \bar{w}} + \tfrac{\delta^2 L}{\delta \bar{w} \delta \bar{w}^T}\tfrac{\delta \bar{w}}{\delta y} = 0
\end{gathered}$$

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 **??** 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 **??**

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

2 Randomised algorithm

Uncertianty based on asymptotic liklihood and $\overline{w}$-distribution
Let $\mathcal{L}_\infty$ be the log-likelihood function for a distribution, now let $\mathcal{L}_N$ denote the log-likelihood function based on N observations from this distribution. Furthermore, let N be a large number, for which $L_N \approx L_\infty$.

$$\mathcal{L}_{\mathcal{N}} = \frac{1}{N}\sum_{n=1}^{N}\ell_n \qquad \overline{w}\, s.t. \frac{\delta \mathcal{L}}{\delta \overline{w}} = \overline{0} \tag{2.1}$$

Where ℓ_n is the log-likelihood of the n^{th} observation. And $\overline{w}$ is the optimal **(true?)** weights for the distribution, then we combine the expressions from (**??**), such that for the optimal weights the following must be fulfilled:

$$\frac{1}{N}\sum_{n=1}^{N}\frac{\delta \ell_n}{\delta \overline{w}} = 0 \tag{2.2}$$

(Skal vi lige skrive lidt om at $\Delta w = w - w_0$ og er en lille forskydelse i vægtene? Eller er det en lille forskydelse?) For each of the N observations, we can write the log-likelihood of the n^{th} observation as:

$$\ell_n(\Delta\overline{w}) = \ell_n\left(\overline{w}_0\right) + \left.\frac{\delta \ell_n}{\delta \overline{w}}\right|_{\overline{w}_0}\Delta\overline{w} + \frac{1}{2}Tr\left[\left.\frac{\delta \ell_n}{\delta \overline{w}\delta \overline{w}^T}\right|_{\overline{w}_0}\Delta\overline{w}\Delta\overline{w}^T\right] \tag{2.3}$$

Or for the entire log-likelihood function:

$$\mathcal{L}_N(\Delta\overline{w}) = \mathcal{L}_\infty\left(\overline{w}_0\right) + \left(\left.\frac{\delta \mathcal{L}_N}{\delta \overline{w}}\right|_{\overline{w}_0}\right)^T \cdot \Delta\overline{w} + \frac{1}{2}\Delta\overline{w}^T\left(\left.\frac{\delta^2 \mathcal{L}_N}{\delta \overline{w}\delta \overline{w}^T}\right|_{\overline{w}_0}\right)\Delta\overline{w} + R \tag{2.4}$$

Where R is the error of the approximation. Furthermore, we define the functions $\overline{\overline{H}}_N = \left.\frac{\delta^2 \mathcal{L}_N}{\delta \overline{w}\delta \overline{w}^T}\right|_{\overline{w}_0}$, and $\overline{g} = \left.\frac{\delta \mathcal{L}_N}{\delta \overline{w}}\right|_{\overline{w}_0}$. And evaluate the condition on $\overline{w}$, stated in (**??**):

$$\frac{\delta \mathcal{L}_{\mathcal{N}}}{\delta \overline{w}} = \overline{g}_N + \overline{\overline{H}}_N\Delta\overline{w} = \overline{0} \tag{2.5}$$

We replace $\Delta\overline{w}$ with $\hat{\Delta\overline{w}}$ as N is a finite number, thus only approximating $\Delta\overline{w}$. Isolating $\hat{\Delta\overline{w}}$, and using Ljung [REFERENCE?], we get:

$$\hat{\Delta\overline{w}} = -\overline{\overline{H}}_N^{-1}\cdot\overline{g}_N \overbrace{=}^{Ljung} -\overline{\overline{H}}_0^{-1}\cdot\overline{g}_{\overline{w}}\left(\overline{w}_0\right) \tag{2.6}$$

(Forklaring af at H_0 er uafhængig af datasæt, mens g nu er afhængig af w evalueret i w_0) Besides getting an estimate for $\Delta\hat{\overline{w}}$, we can find the mean of the distribution:

$$\left\langle \Delta\hat{\overline{w}} \right\rangle = -\overline{\overline{H}}_0^{-1} \left\langle \overline{g}_w \right\rangle (\overline{w}_0)$$

Nu skal vi prøve at kombinere flere datasets