Stochastic Gradient Descent in NPIV estimation

Anonymous Author(s)
Affiliation
Address
email

1 Binary response models

We want to be able to employ the same risk minimization procedure:

$$\arg\min_{h\in\mathcal{F}} \mathcal{R}(h) = \arg\min_{h\in\mathcal{F}} \mathbb{E}_Z \left[\ell(r_0(Z), \mathcal{P}[h](Z))\right]. \tag{1}$$

Let's see what data generating procedure makes this possible. Firstly, let

$$Y \mid X, \varepsilon \sim \mathrm{Bernoulli}(\sigma(h^\star(X) + \varepsilon)), \tag{2}$$

where σ is the logistic function, $\mathbb{E}[\varepsilon \mid X] \neq 0$ and $\mathbb{E}[\varepsilon \mid Z] = 0$. For (1) to make sense, we'd like $r_0(Z) = \mathbb{E}[Y \mid Z]$ and $\mathcal{P}[h^\star](Z) = \mathbb{E}[h^\star(X) \mid Z]$ to be close according to a suitable loss function ℓ, at least close enough so that $h^\star$ is a solution to (1). Let's see if this is the case under (2):

$$\mathbb{E}[Y \mid Z] = \mathbb{P}[Y = 1 \mid Z],$$

Assuming (2), we may compute this conditioning on X and ε and then integrating them out:

$$\begin{aligned}
\mathbb{P}[Y = 1 \mid Z = z] &= \int_{\mathcal{X}\times\mathbf{R}} \mathbb{P}[Y = 1 \mid Z = z, X = x, \varepsilon = e] p_{X,\varepsilon|Z}(x, e \mid z)\, \mathrm{d}x\mathrm{d}\varepsilon \\
&= \int_{\mathcal{X}\times\mathbf{R}} \sigma(h^\star(x) + e) p_{X,\varepsilon|Z}(x, e \mid z)\, \mathrm{d}x\mathrm{d}\varepsilon \\
&= \mathbb{E}[\sigma(h^\star(X) + \varepsilon) \mid Z = z].
\end{aligned}$$

There are a two main problems here. The first one is that ε appears inside σ and, hence, does not vanish after conditioning on $Z = z$. I cannot think of a way to remove it without assuming known the distribution of ε given X, which is prohibitive. The second problem is that, even if there was no ε, the expectation is outside the function σ. In order for (1) to work under (2), we'd like set

$$\ell(y, y') = \mathrm{BCE}(y, \sigma(y')),$$

where BCE is the binary cross entropy loss function:

$$\mathrm{BCE}(y, p) = -[y \log p + (1 - y) \log(1 - p)].$$

That is, we'd like to have $\sigma\left(\mathbb{E}[h(X) \mid Z]\right)$ inside $\mathcal{R}(h)$, instead of $\mathbb{E}\left[\sigma\left(h(X)\right) \mid Z\right]$.

The second option is to set

$$Y = \mathbf{1}[h^\star(X) + \varepsilon > 0]. \tag{3}$$

Here, we have

$$\mathbb{E}[Y \mid Z = z] = \mathbb{P}[h^\star(X) + \varepsilon > 0 \mid Z = z]. \tag{4}$$

To try to make this lead somewhere, let's define $\eta = h^\star(X) - \mathbb{E}[h^\star(X) \mid Z] + \varepsilon$, so that

$$Y = \mathbf{1}[\mathbb{E}[h^\star(X) \mid Z] + \eta > 0]$$

and $\mathbb{E}[\eta \mid Z] = 0$. Let $t(Z) = \mathbb{E}[h^\star(X) \mid Z]$. This implies

$$\begin{aligned}\mathbb{E}[Y \mid Z] &= \mathbb{P}[t(z) + \eta > 0 \mid Z] \\ &= 1 - F_{\eta|Z}(-t(Z)).\end{aligned}$$

Hence, we have

$$t(Z) = -F^{-1}_{\eta|Z}(r_0(Z) - 1).$$

Or, equivalently:

$$\mathbb{E}[h^\star(X) \mid Z] = -F^{-1}_{\eta|Z}\left(\mathbb{E}[Y \mid Z] - 1\right).$$

This looks promising: If we assume to know the conditional distribution of η given Z, we have a couple of options. We can minimize

$$\mathrm{BCE}(r_0(Z), 1 - F_{\eta|Z}(-\mathbb{E}[h^\star(X) \mid Z])),$$

or

$$\left(\mathbb{E}[h^\star(X) \mid Z] + F^{-1}_{\eta|Z}(r_0(Z) - 1)\right)^2.$$

This assumption was used on the paper “Nonparametric Instrumental Variable Estimation of Binary Response Models”, by P. L. Florens, from where I took the ideas for these calculations.