SVM conditional probability estimation (linear vs nonlinear)

Chanwoo Lee, April 16, 2020

1 Model setting

Feature and label data set is given as $(\boldsymbol{x}_1, y_1), \ldots, (\boldsymbol{x}_N, y_N)$ where $\boldsymbol{x}_i \in \mathbb{R}^2$ and $y_i \in \{-1, 1\}$ for $i = 1, \ldots, N$. We generate the data set with the conditional probability,

$$\boldsymbol{x}|y = 1 \sim N(\boldsymbol{\mu}_1, 1) \quad \text{and} \quad \boldsymbol{x}|y = -1 \sim N(\boldsymbol{\mu}_2, 1).$$

We can calculate the conditional probability $\mathbb{P}(y = 1|\boldsymbol{x})$ with the assumption that $\mathbb{P}(y = 1) = \mathbb{P}(y = -1)$ as

$$\begin{aligned}
\mathbb{P}(y = 1|\boldsymbol{x}) &= \frac{\mathbb{P}(\boldsymbol{x}|y = 1)\mathbb{P}(y = 1)}{\mathbb{P}(\boldsymbol{x}|y = -1)\mathbb{P}(y = -1)} \qquad (1)\\
&= \frac{\exp\left(-\frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu}_1)^T(\boldsymbol{x} - \boldsymbol{\mu}_1)\right)}{\exp\left(-\frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu}_1)^T(\boldsymbol{x} - \boldsymbol{\mu}_1)\right) + \exp\left(-\frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu}_2)^T(\boldsymbol{x} - \boldsymbol{\mu}_2)\right)}\\
&= \frac{1}{1 + \exp\left((\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1)(x - \frac{\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2}{2})\right)}.
\end{aligned}$$

Therefore, our objective function value we want to estimate for each given π is that

$$\begin{aligned}
\text{sign}\left(\mathbb{P}(y = 1|\boldsymbol{x}) - \pi\right) &= \text{sign}\left(\frac{1}{1 + \exp\left((\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1)(x - \frac{\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2}{2})\right)} - \pi\right) \qquad (2)\\
&= \text{sign}\left((\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1)^T\boldsymbol{x} - \frac{\|\boldsymbol{\mu}_1\|^2 - \|\boldsymbol{\mu}_2\|^2}{2} - \log\left(\frac{1 - \pi}{\pi}\right)\right)\\
&= \text{sign}\left(\boldsymbol{w}^T\boldsymbol{x} + b\right),
\end{aligned}$$

where $\boldsymbol{w} = \boldsymbol{\mu}_2 - \boldsymbol{\mu}_1$ and $b = -\frac{\|\boldsymbol{\mu}_1\|^2 - \|\boldsymbol{\mu}_2\|^2}{2} - \log\left(\frac{1-\pi}{\pi}\right)$.

2 Simulation result

With the same setting in the precvious section, we perform simulation with $N = 25, \boldsymbol{\mu}_1 = (0, 0)^T, \boldsymbol{\mu}_2 = (1.5, 1.5)^T$ and $\#\{i : y_i = 1\}$ is 12. Figure 1 shows the estimated conditional probability $\mathbb{P}(y = 1|\boldsymbol{x})$ with linear SVM and nonlinear SVM (Gaussian). In this case, we can check linear kernel outperforms the Gaussian kernel considering ground truth probability we calculated in (1). The possible reason for this is that our true objective function in (2) can be written as linear function of the feature data $\boldsymbol{x}$. This is why the linear model estimation is better than non linear model.

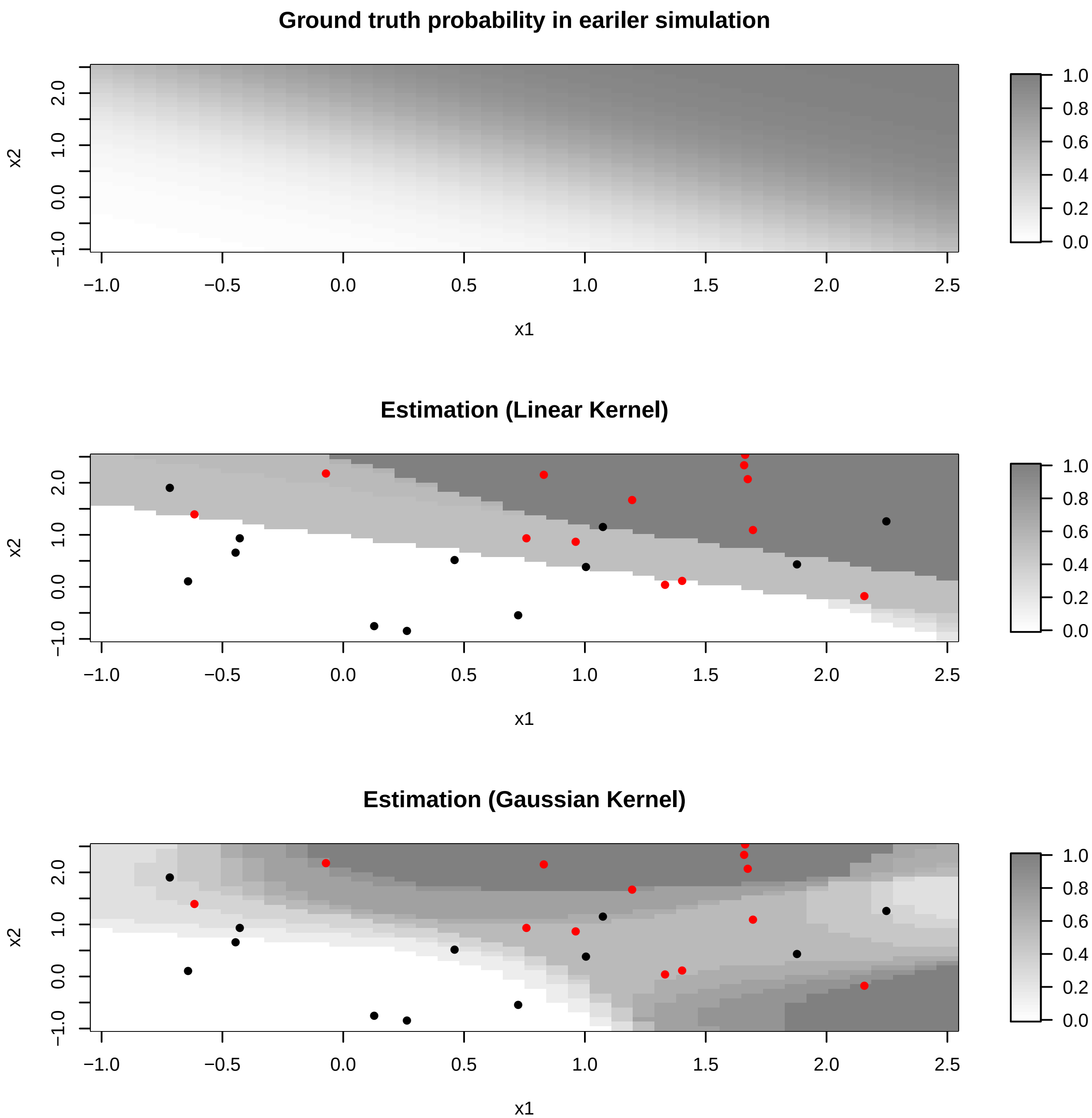

Figure 1: The first sub figure shows the true conditional probability $\mathbb{P}(y = 1|\boldsymbol{x})$. The second and third one show estimated conditional probability with linear kernel and Gaussian kernel respectively.