Conditional Gamma Distribution: Bond Balance Prediction Problem

1 Problem setup

- Input space: $\mathcal{X} = \mathbf{R}^d$
- Outcome space: $\mathcal{Y} = \{y \in \mathbf{R} \mid y \geq 0\}$
- Action space: Distributions on $\mathcal{Y}$.

2 Modeling Decisions

We went to google to find a family of densities that has the right support ($\mathcal{Y}$) and seems appropriate for the problem. We came up with the family of Gamma distributions with shape parameter $\theta = 1$. The density is then

$$p(y \mid k) = \frac{1}{\Gamma(k)} y^{k-1} e^{-y},$$

for parameter $k \in (0, \infty)$. Support for this density is $y \in (0, \infty)$.

We want to find a prediction function $f : x \mapsto k$, where k is the parameter of our parametric family. Once we have k, the final probability distribution produced is the Gamma distribution with parameter $k = f(x)$ and $\theta = 1$.

We will use a linear model, in the sense that all information we are extracting from x can be summarized by a single linear function of x. From there, we'll need to produce the parameter estimate k. So, introducing $w \in \mathbf{R}^d$ to give us the linear function, and write

$$x \mapsto \underbrace{w^T x}_{\mathbf{R}} \mapsto \underbrace{\sigma(w^T x)}_{(0,\infty)} = k,$$

for some transfer function $\sigma : \mathbf{R} \to \mathbf{R}^{>0}$, which we still need to determine. Remember the transfer function maps us from output of our linear function, which can be anything in $\mathbf{R}$, to our parameter space, which is $(0, \infty)$.

How about

$$\sigma(s) = e^s.$$

So the final prediction function is

$$f(x; w) = \exp(w^T x).$$

3 Model Fitting

Our final prediction function is $f(x; w) = \exp(w^T x)$, but we still need to choose $w \in \mathbf{R}^d$.

Suppose we have a training sample

$$\mathcal{D} = ((x_1, y_1), \ldots, (x_n, y_n))$$

sampled i.i.d. (so the y_i's are independent given the x's) where $(x_i, y_i) \in \mathbf{R}^d \times (0, \infty)$.

We'll follow the maximum likelihood approach. We start by writing down the likelihood function, which gives us the density for $\mathcal{D}$ for any w:

$$\begin{aligned}
L_{\mathcal{D}}(w) &= \prod_{i=1}^{n} p(y_i \mid x_i, w) \\
&= \prod_{i=1}^{n} \frac{1}{\Gamma(\exp(w^T x_i))} y_i^{\exp(w^T x_i) - 1} e^{-y_i}.
\end{aligned}$$

It will be convenient to compute the log of this, so start with

$$\begin{aligned}
\log p(y_i \mid x_i, w) &= \log \left[\frac{1}{\Gamma(\exp(w^T x_i))} y_i^{\exp(w^T x_i) - 1} e^{-y_i} \right] \\
&= \log \left[\frac{1}{\Gamma(\exp(w^T x_i))} \right] \\
&\quad + \log \left[y_i^{\exp(w^T x_i) - 1} \right] - y_i \\
&= -\log \left[\Gamma(\exp(w^T x_i)) \right] \\
&\quad + \left[\exp(w^T x_i) - 1 \right] \log y_i - y_i
\end{aligned}$$

Following the approach of maximum likelihood, let's choose w to maximize $L_{\mathcal{D}}(w)$. Equivalently, let's maximize the log-likelihood. So

$$w^*_{\text{MLE}} = \underset{w \in \mathbf{R}^d}{\arg\max} \log L_{\mathcal{D}}(w)$$

where

$$\begin{aligned} \log L_{\mathcal{D}}(w) &= \sum_{i=1}^{n} [-\log\left[\Gamma(\exp(w^T x_i))\right] \\ &\quad + \left[\exp(w^T x_i) - 1\right] \log y_i - y_i] \end{aligned}$$

Equivalent to find

$$\underset{w \in \mathbf{R}^d}{\arg\max} \sum_{i=1}^{n} [-\log\left[\Gamma(\exp(w^T x_i)) + \exp(w^T x_i) \log y_i\right]$$

So we just need to optimize this over w, and we've got our prediction functions.

In the future, we'll learn how to swap out the linear piece $w^T x$ with something nonlinear, such as a gradient boosted regression tree model or a neural network.