Overview of models

20 Sep 2023

This is an attempt

Linear models

$$y_i = (\mathbf{X}\beta)_i + \epsilon_i, \;\; \epsilon \sim \text{Normal}(0, \sigma^2)$$

or

$$y_i \sim \text{Normal}((\mathbf{X}\beta)_i, \sigma^2)$$

(this form generalizes better to response distributions where we can't shift the location by adding a term)

or

$$\mathbf{y} \sim \text{MVN}(\mathbf{X}\beta, \sigma^2 I)$$

Generalized linear models

As above, but add a monotonic, pre-determined (no free parameters) *link function* f and a distribution Dist from the **exponential family**. Then

$$y_i \sim \text{Dist}(f^{-1}(\mathbf{X}\beta)_i, \phi)$$

where ϕ is a **scale parameter**.

In the case where Dist is Gaussian, f is the identity, and $\phi = \sigma^2$, this reduces to the linear model. When (for example) Dist is Bernoulli, f is $\log(p/(1-p)$ (the *logit* or log-odds function), and $\phi = 1$, this is **logistic regression**.

Additive models

Make **X** a piecewise polynomial basis with continuous derivatives, most often cubic. There are lots of ways to set up such bases.

Ridge regression

There are a variety of ways to set this up. The most common is as a **penalized** regression, i.e. say that we want

$$\arg\min \beta ||(\mathbf{X}\beta - \mathbf{y})||_2^2 + \lambda ||\beta||_2^2$$

i.e., we want to minimize the sum of squared deviations of the regression model from the data, plus the sum of squared beta values, with a penalty weight of λ. We could equivalently set this up as a likelihood problem: find the MLE of

$$\int L(y|\beta, \sigma_r^2) \cdot L(\beta|\sigma_g^2)\, d\beta$$

where we assume that $y_i \sim \text{Normal}((\mathbf{X}\beta)_i, \sigma_r^2)$ and $\beta_i \sim \text{Normal}(0, \sigma_g^2)$.

(The integral disappears for linear mixed models.)

This is also equivalent to a Bayesian model where we impose iid Normal zero-centred priors on the elements of β (to make it fully Bayesian, we would need to specify priors for σ_r^2 and $□^2_g).

Mixed models

Similar, but instead of putting the penalty on the regression parameters (or equivalently treating the regression parameters as having , we will put the priors on **random effects** parameters that describe the deviation of cluster-level values from population values.

The simplest case (described in an R formula as `y ~ 1 + (1|g)`) is a model with a population-level intercept β_0 and group-level deviations from the population mean b_i.

This case, and more complex cases, can be written as

$$\begin{aligned} y_i &\sim \text{Normal}((\mathbf{X}\beta + \mathbf{Zb})_i, \sigma_r^2) \\ \mathbf{b} &\sim \text{MVN}(0, \Sigma(\theta)) \end{aligned}$$

where θ is a vector of parameters that defines the covariance matrix Σ.

Generalized linear mixed models

The same, but add a link function and an exponential-family distribution.

Generalized additive mixed models

The same, but allow the parameters describing the spline (or whatever) basis to be penalized/shrunk toward zero.

References