



THE UNIVERSITY OF TEXAS AT AUSTIN

Department of Statistics and Data Sciences

College of Natural Sciences

Putting it all together

Sinead Williamson

The University of Texas Department of Statistics and Data Science

Starting point: Bayesian linear regression

Basic model:

$$\mathbf{y}|\beta, X \sim \text{Normal}(X\beta, (\omega\Lambda)^{-1})$$

$$\beta \sim \text{Normal}(\mu, (\omega K)^{-1})$$

$$\omega \sim \text{Gamma}(a, b)$$

Let's look at what this looks like... [notebook]

We can modify this in a wide variety of ways! Some of which we've played around with...

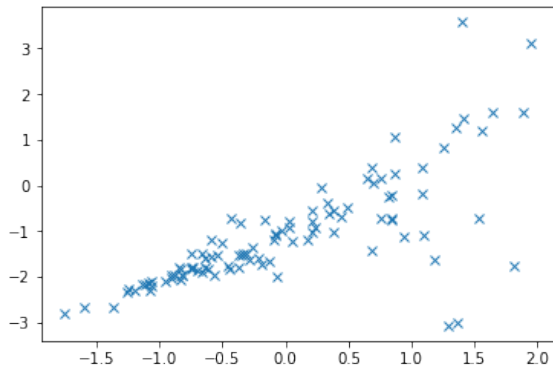
- ▶ Switch out likelihood for a different distribution
- ▶ Allow variance to vary between individuals → heavy tails
- ▶ Allow variance to vary between groups

We can modify this in a wide variety of ways! Some of which we've played around with...

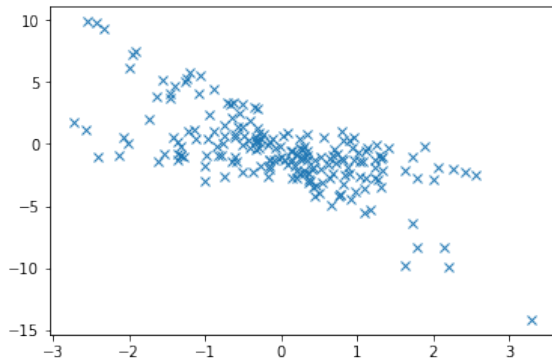
- ▶ Switch out likelihood for a different distribution
- ▶ Allow variance to vary between individuals → heavy tails
- ▶ Allow variance to vary between groups

Some of which we haven't!

How could we model data that looks like this?



How about data that looks like this?



What could we do if we were missing covariates?

X1	X2	X3	Y
1.45	0.22	0.73	3.88
0.62	-	1.21	1.56
2.21	1.67	1.08	3.42

The extreme version of missing (categorical) covariates is a Gaussian mixture model:

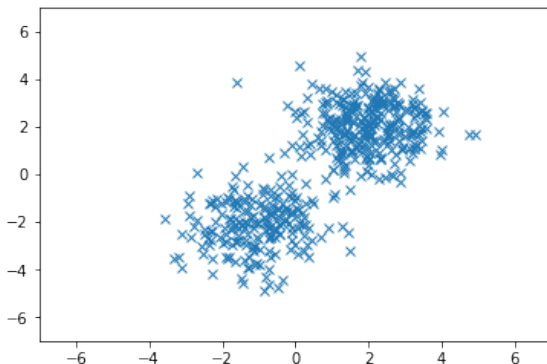
$$\pi \sim \text{Dirichlet}(\alpha)$$

$$Z_i \sim \pi$$

$$\mu_k \sim \text{Normal}(\mu_0, \sigma_0^2)$$

$$\omega_k \sim \text{Gamma}(a, b)$$

$$X_i \sim \text{Normal}(\mu_{Z_k}, 1/\omega_{Z_k})$$



$$\pi \sim \text{Dirichlet}(\alpha) \quad Z_i \sim \pi$$

$$\mu_k \sim \text{Normal}(\mu_0, \sigma_0^2) \quad \omega_k \sim \text{Gamma}(a, b) \quad X_i \sim \text{Normal}(\mu_{Z_k}, 1/\omega_{Z_k})$$

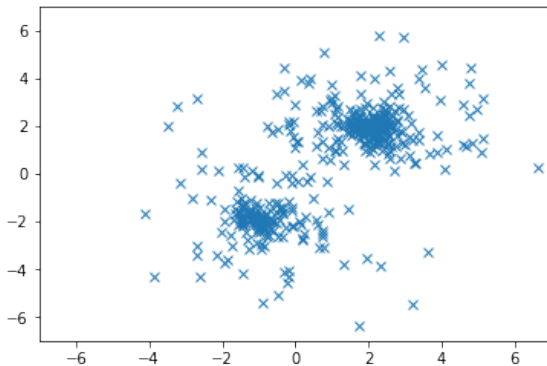
- ▶ Conditioned on Z_i , we have a linear regression model. We can either sample ω and μ , or integrate them out.
- ▶ Conditioned on π , ω and μ , we have

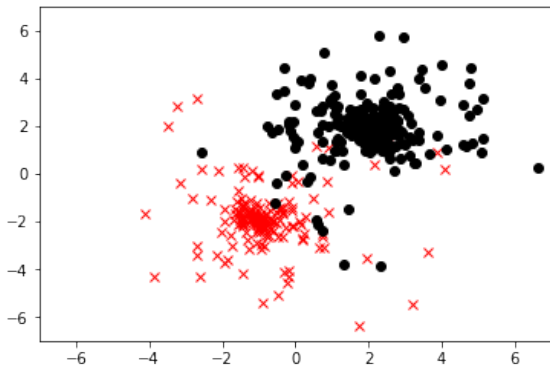
$$P(Z_i = k | \theta, \mu, \omega) \propto \pi_k N(X_i; \mu_k, 1/\omega_k)$$

- ▶ Can construct a normalized vector $\hat{p} : \hat{p}_k = \frac{\pi_k N(X_i; \mu_k, 1/\omega_k)}{\sum_j \pi_j N(X_i; \mu_j, 1/\omega_j)}$
- ▶ Can then sample from a multinomial with probability \hat{p} .
- ▶ Can integrate out π :

$$P(Z_i = k | \theta, \mu, \omega) \propto \left(\sum_{j \neq i} I(Z_j = k) \right) N(X_i; \mu_k, 1/\omega_k)$$

What could we do if our data looked like this?





If we get a mixture model by putting a prior over latent categorical regressors...

What do we get if we put a prior over latent continuous valued regressors?