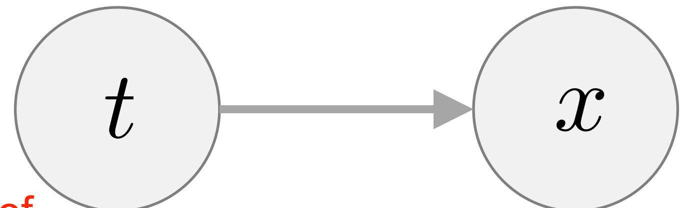
Introducing latent variable

$$p(x \mid \theta) = \pi_1 \mathcal{N}(x|\mu_1, \Sigma_1) + \pi_2 \mathcal{N}(x|\mu_2, \Sigma_2) + \pi_3 \mathcal{N}(x|\mu_3, \Sigma_3)$$

Now, suppose we alter our GMM model from previous lessons as follows:
Suppose that each variable x was generated using info from its own latent variable t.
Every t explains an x.



t here takes 3 values {1,2,3} (# of classes)

It shows us from which gaussian each datapoint came from. But we don't know this, hence it is 'latent'.

Somewhat like the latent 'label'.

$$p(t = c \mid \theta) = \pi_c$$

< this is the prior distribution of latent var t. Exactly same as weight of gaussians.

$$p(x \mid t = c, \theta) = \mathcal{N}(x \mid \mu_c, \Sigma_c)$$

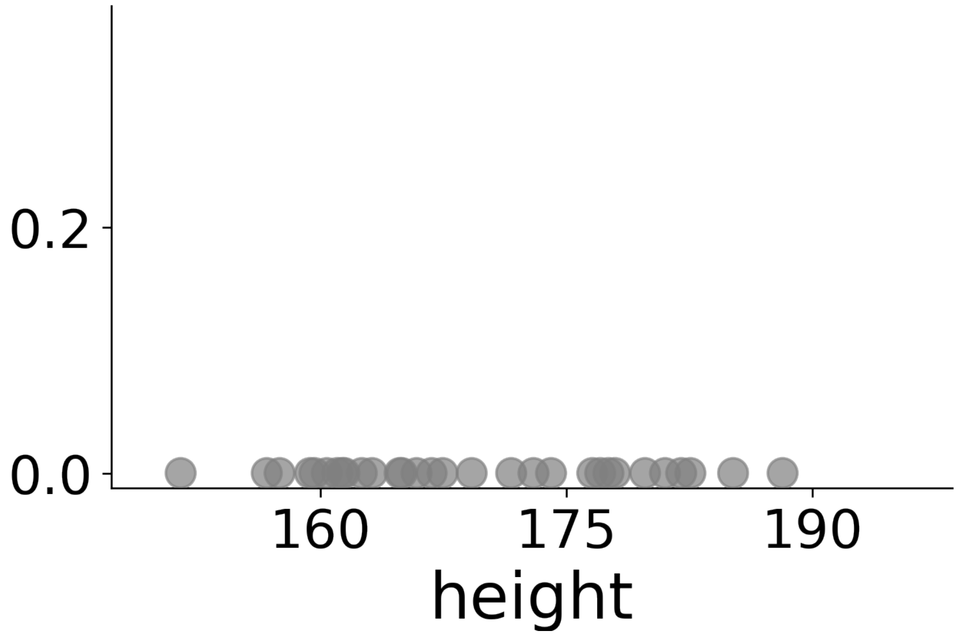
< The likelihood, (density of point x given t=c, and theta) is just the c gaussian.

$$p(x \mid \theta) = \sum_{c=1}^{3} p(x \mid t = c, \theta) p(t = c \mid \theta)$$

We marginalize over the t values to get p(x|theta) (sum rule)

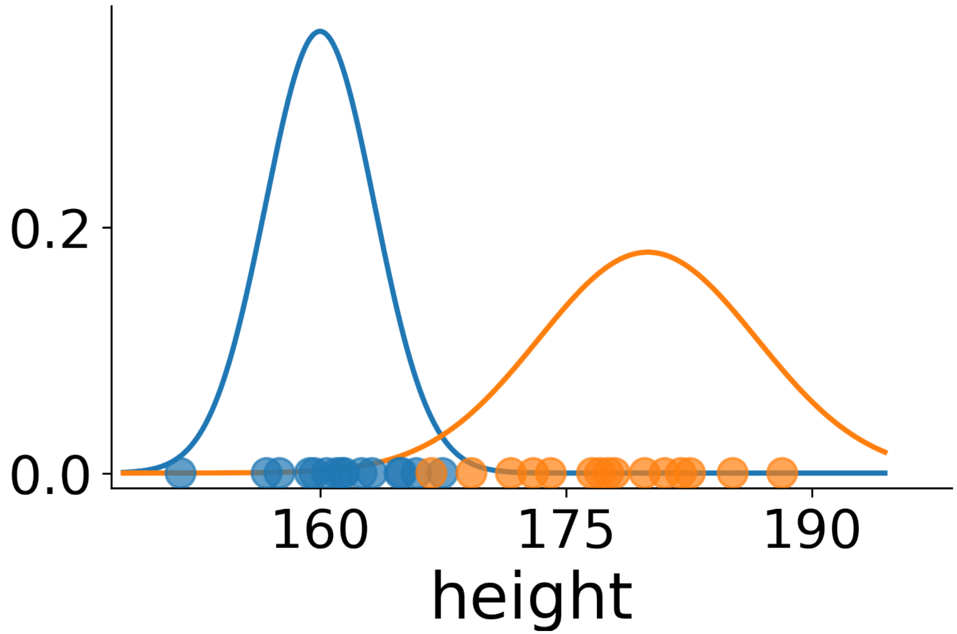
NOTE: I should revise more about bayesian marginalization. Not sure why this is correct ^^

Expectation Maximization



Dataset: $\{x_1, \ldots, x_N\}$

Expectation Maximization



How to estimate parameter θ?

If sources t are known, easy:

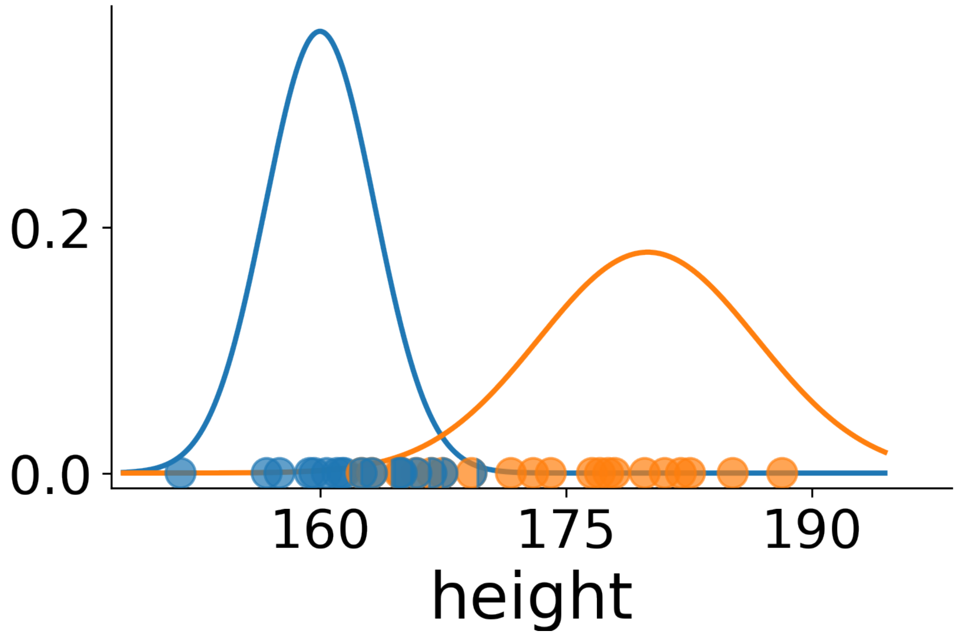
ALSO if we know where each point belongs to:

$$p(x \mid t = 1, \theta) = \mathcal{N}(x \mid \mu_1, \sigma_1^2)$$

$$\mu_1 = \frac{\sum_{\text{blue } i} x_i}{\text{\# of blue points}} \qquad \sigma_1^2 = \frac{\sum_{\text{blue } i} (x_i - \mu_1)^2}{\text{\# of blue points}}$$

Simply get the mean of the data points, and for variance, just sum of squared difference, divided by # of blue points (we suppose that each point has 1/N probability of showing up here).

Expectation Maximization



How to estimate parameter θ?

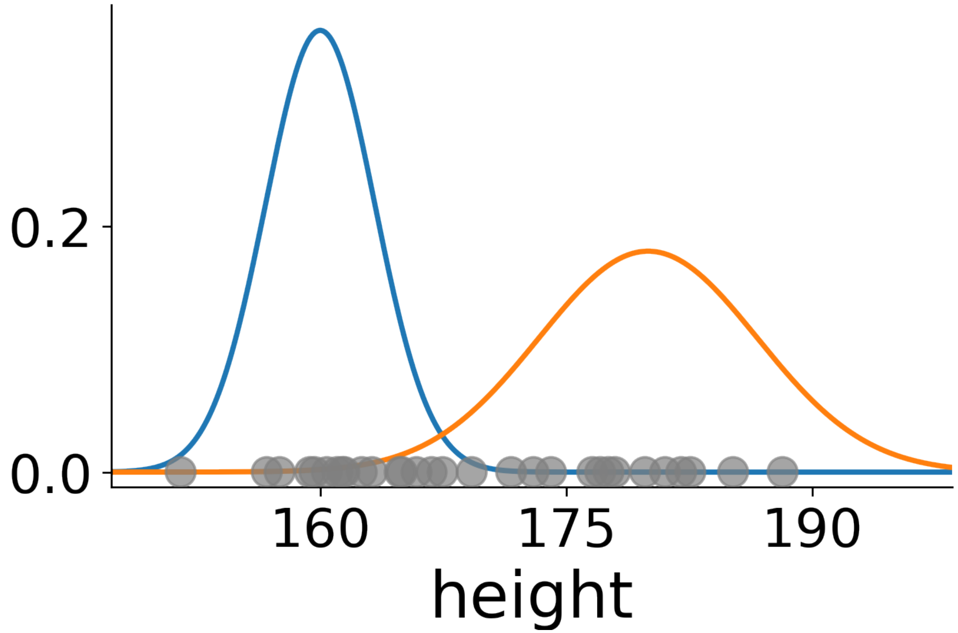
If we don't know where each point belongs too, but we know each t's distributions:

If sources t are known, easy:

$$\mu_1 = \frac{\sum_i p(t_i = 1 \mid x_i, \theta)\, x_i}{\sum_i p(t_i = 1 \mid x_i, \theta)} \quad \sigma_1^2 = \frac{\sum_i p(t_i = 1 \mid x_i, \theta)\, (x_i - \mu_1)^2}{\sum_i p(t_i = 1 \mid x_i, \theta)}$$

blue is t_i = 1.

Expectation Maximization



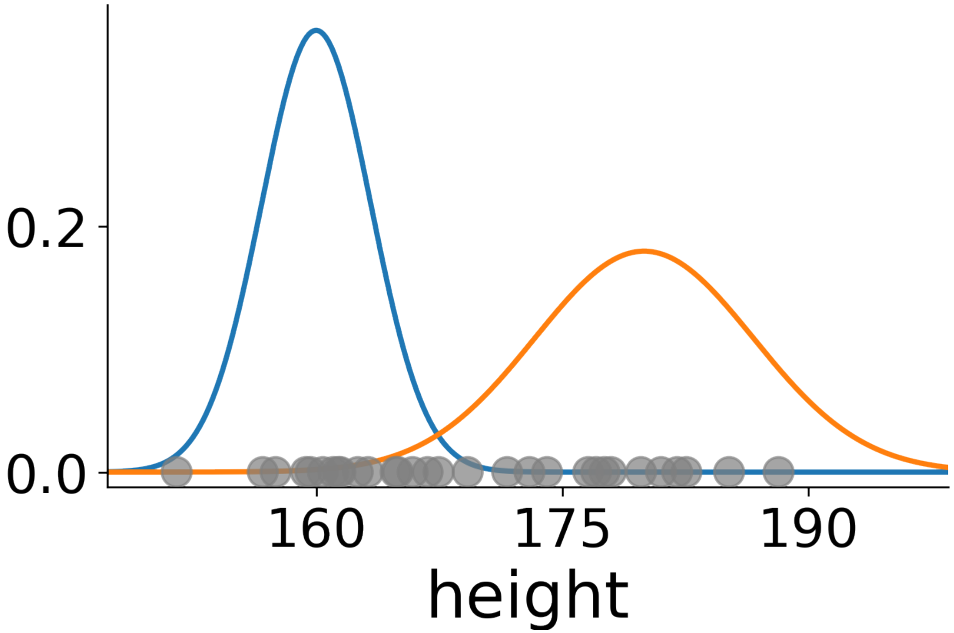
What if we don't know the sources?

In unsupervised learning, we often don't KNOW the sources!

Given: $$p(x \mid t = 1, \theta) = \mathcal{N}(-2, 1)$$

Find: $$p(t = 1 \mid x, \theta)$$

Expectation Maximization



What if we don't know the sources?

If we know parameters θ, easy:

$$p(t = 1 \mid x, \theta) = \frac{p(x \mid t = 1, \theta)\, p(t = 1 \mid \theta)}{Z}$$

Here, Z is the normalization constant.

Expectation Maximization

Chicken and egg problem

- Need Gaussian parameters to estimate sources
- Need sources to estimate Gaussian parameters

EM algorithm

1. Start with 2 randomly placed Gaussians parameters θ
2. Until convergence repeat:
   a) For each point compute $p(t = c \mid x_i, \theta)$: does x_i look like it came from cluster c?
   b) Update Gaussian parameters θ to fit points assigned to them