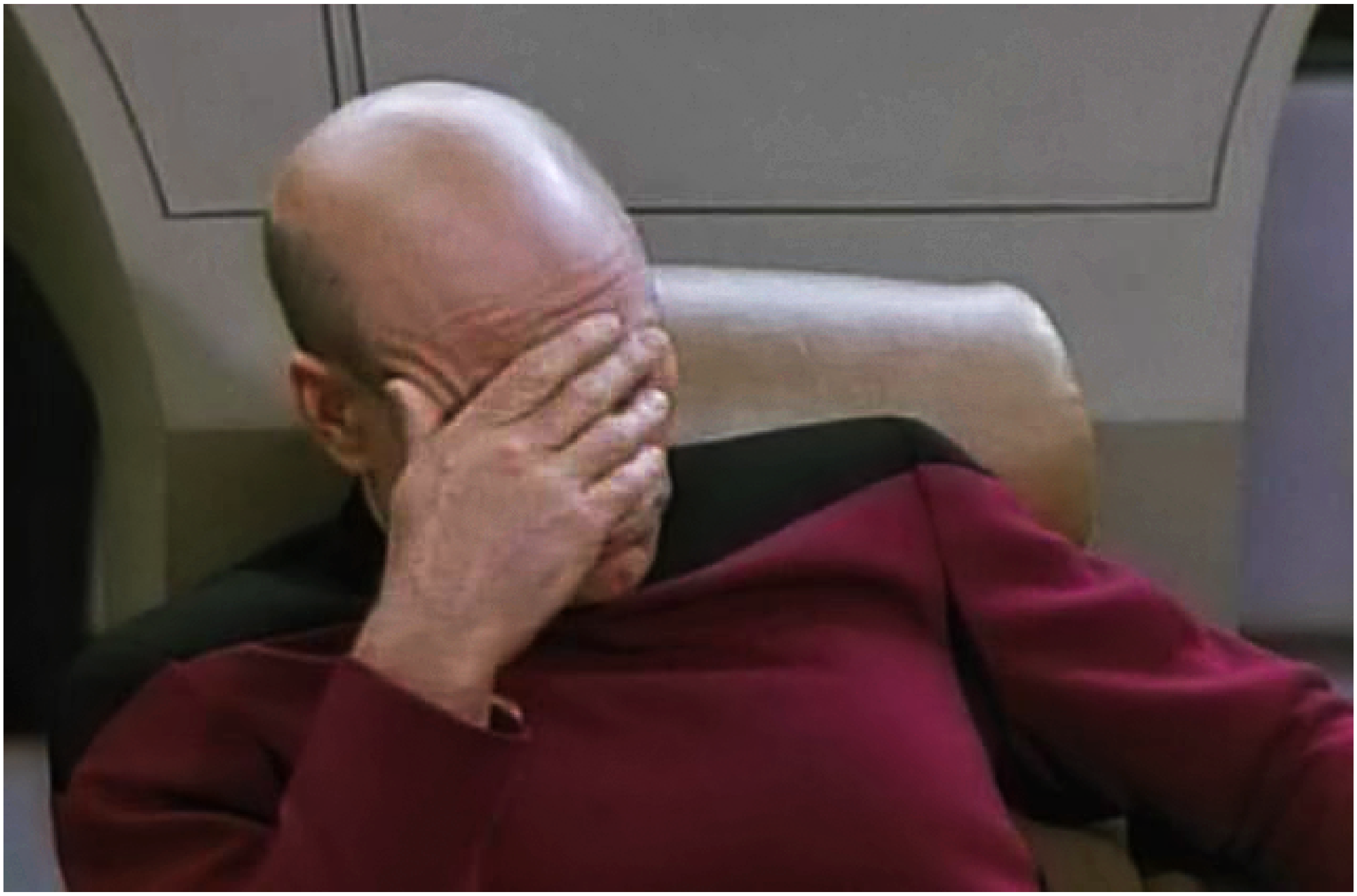

В конце лекции я написал формулу:

$$Q(\theta, \theta_{old}) = \sum_{i=1}^{n} P(z_i = 1 \mid x, \theta_{old}) \left(-\frac{1}{2} \ln(2\pi\sigma_1^2) - \frac{1}{2} \frac{(x_i - \mu_1)^2}{\sigma_1^2} \right) + (1 - P(z_i = 1 \mid x, \theta_{old})) \left(-\frac{1}{2} \ln(2\pi\sigma_2^2) - \frac{1}{2} \frac{(x_i - \mu_2)^2}{\sigma_2^2} \right) \quad (1)$$

Это неверно! Снимаем лапшу с ушей!

Так как нам нужен логарифм от совместной плотности в ожидании $E(\ln p(x, z|\theta) \mid x, \theta_{old})$, то нужно добавить $\ln P(z_i = 1|\theta)$ и $\ln(1 - P(z_i = 1|\theta))$. Плотность домножается, поэтому логарифм прибавляем.

$$Q(\theta, \theta_{old}) = \sum_{i=1}^{n} P(z_i = 1 \mid x, \theta_{old}) \left(-\frac{1}{2} \ln(2\pi\sigma_1^2) - \frac{1}{2} \frac{(x_i - \mu_1)^2}{\sigma_1^2} + \ln p_1 \right) + (1 - P(z_i = 1 \mid x, \theta_{old})) \left(-\frac{1}{2} \ln(2\pi\sigma_2^2) - \frac{1}{2} \frac{(x_i - \mu_2)^2}{\sigma_2^2} + \ln(1 - p_1) \right) \quad (2)$$

В результате оптимальные μ_1^{new}, μ_2^{new}, $\sigma_{1,new}^2$, $\sigma_{2,new}^2$ не меняются! Это меня и смутило, я ожидал, что при ошибке в формуле поменяются оптимумы.

А вот оптимальное p_1^{new} оказывается результатом честной оптимизации, а не эвристики, как я сказал.