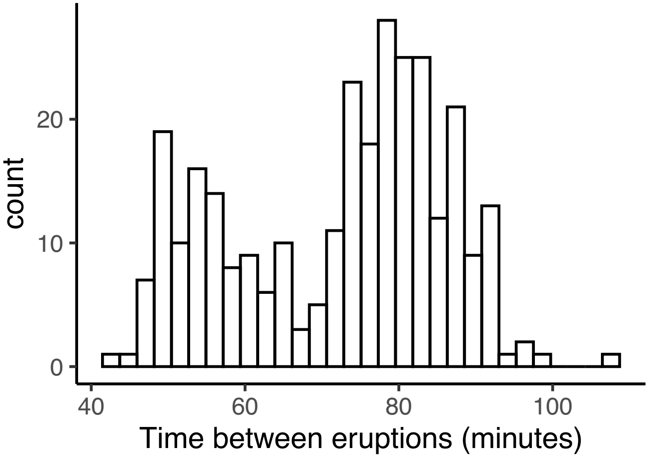
Lecture 31: Gaussian mixture models with multivariate data

Ciaran Evans

Course logistics

- HW 8 released, due Monday 4/28
- This week: wrap up Gaussian mixtures and EM algorithm
- Course evals: Wednesday during class
- Monday 4/28: wrap-up work day

Previously



Previously: Gaussian mixture model

- Observe data $X_1, ..., X_n$
- Assume each observation i comes from one of k groups. Let $Z_i \in \{1, ..., k\}$ denote the group assignment
  - The group Z is an unobserved (**latent**) variable

Model:

- $P(Z_i = j) = \lambda_j$
- $X_i | (Z_i = j) \sim N(\mu_j, \sigma_j^2)$

Posterior probabilities and parameter estimation

- **If** we know the parameters λ, μ, σ, we can calculate posterior probabilities:

$$P(Z_i = j|X_i) = \frac{\lambda_j f(X_i|Z_i = j)}{\lambda_1 f(X_i|Z_i = 1) + \cdots + \lambda_k f(X_i|Z_i = k)}$$

- **If** we know the posterior probabilities, we can estimate the model parameters λ, μ, and σ:
  - $\widehat{\lambda}_j = \frac{1}{n} \sum\limits_{i=1}^{n} P(Z_i = j|X_i)$
  - $\widehat{\mu}_j = \dfrac{\sum\limits_{i=1}^{n} X_i P(Z_i = j|X_i)}{\sum\limits_{i=1}^{n} P(Z_i = j|X_i)}$
  - $\widehat{\sigma}_j = \sqrt{\dfrac{\sum\limits_{i=1}^{n} (X_i - \widehat{\mu}_j)^2 P(Z_i = j|X_i)}{\sum\limits_{i=1}^{n} P(Z_i = j|X_i)}}$

Putting everything together

Model: $P(Z_i = j) = \lambda_j \qquad X_i | (Z_i = j) \sim N(\mu_j, \sigma_j^2)$

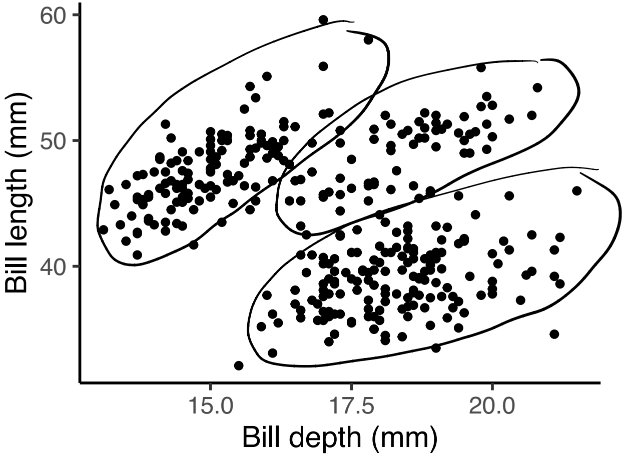
Parameters: $\lambda = (\lambda_1, ..., \lambda_k)$, $\mu = (\mu_1, ..., \mu_k)$, $\sigma = (\sigma_1, ..., \sigma_k)$

Estimation:

1. Initialize parameter guesses $\lambda^{(0)}$, $\mu^{(0)}$, $\sigma^{(0)}$
2. Given current parameter estimates, compute $P^{(0)}(Z_i = j | X_i)$ for all i, j
3. Given current posterior probabilities $P^{(0)}(Z_i = j | X_i)$, update parameter estimates to $\lambda^{(1)}$, $\mu^{(1)}$, $\sigma^{(1)}$
4. Iterate: repeat steps 2–3 until convergence

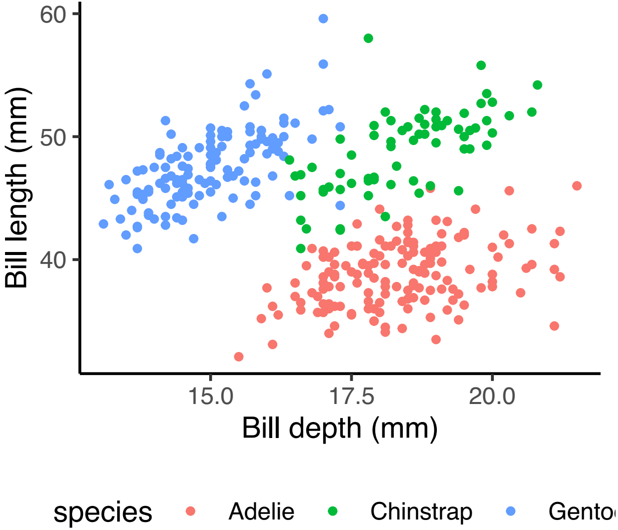
Multivariate data

Penguin data:

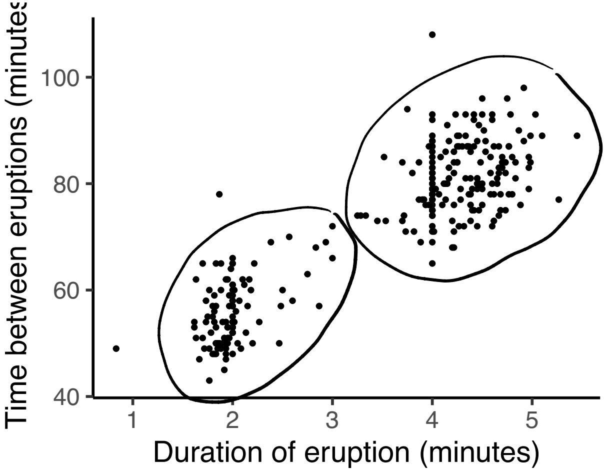


Question: What do you notice about this scatterplot?

Multivariate data



Multivariate data



Question: How should we generalize our Gaussian mixture model to multivariate data?

Multivariate normal distribution

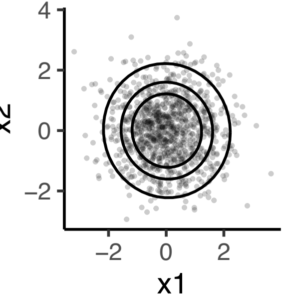
Definition: Let $X = (X_1, ..., X_k)^T$. We say that $X \sim N(\mu, \Sigma)$ if for any $\mathbf{a} \in \mathbb{R}^k$, $\mathbf{a}^T X$ follows a (univariate) normal distribution.

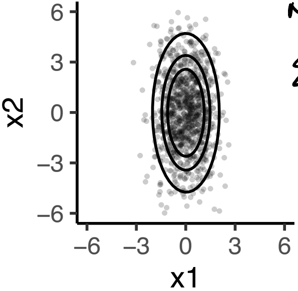
- $\mu = \mathbb{E}[X] = \begin{bmatrix} \mathbb{E}[X_1] \\ \mathbb{E}[X_2] \\ \vdots \\ \mathbb{E}[X_k] \end{bmatrix}$ (coordinate-wise expectations)
- $\Sigma = Var(X) = \begin{bmatrix} Var(X_1) & Cov(X_1, X_2) & \cdots & Cov(X_1, X_k) \\ Cov(X_2, X_1) & Var(X_2) & & \\ & & \ddots & \\ & & & Var(X_k) \end{bmatrix}$ (variance-covariance matrix)

Multivariate normal distribution

$N(0, \Sigma)$

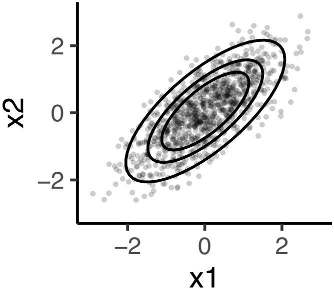
$\Sigma = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$





$N(0, \Sigma)$

$\Sigma = \begin{pmatrix} 1 & 0 \\ 0 & 5 \end{pmatrix}$



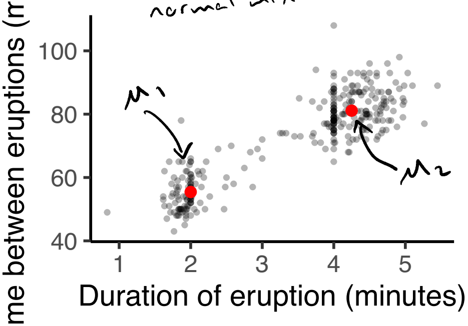
$N(0, \Sigma)$

$\Sigma = \begin{pmatrix} 1 & 0.75 \\ 0.75 & 1 \end{pmatrix}$

Multivariate Gaussian mixture model

```r
em_res <- mvnormalmixEM(old_faithful,
                        lambda = c(0.5, 0.5), k=2)
```

multivariate normal mixture



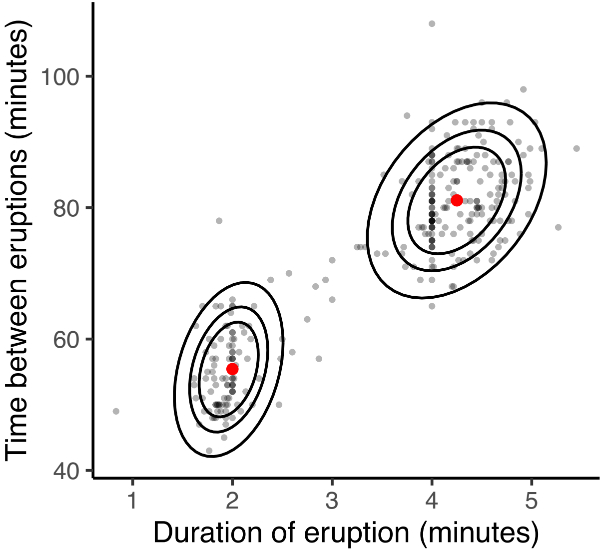

```r
em_res$mu[[1]]
```

```
## [1]  1.966059 55.430118
```

```r
em_res$mu[[2]]
```

```
## [1]  4.250835 81.114544
```

Multivariate Gaussian mixture model



Your turn

Implement the algorithm to fit a Gaussian mixture model:

https://sta379-s25.github.io/practice_questions/pq_31.html

- Start in class
- Welcome to work with a neighbor
- Solutions are posted on the course website