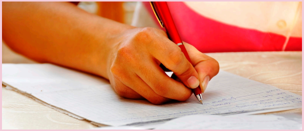
Exercise: Multivariate Normal model

Erik Štrumbelj

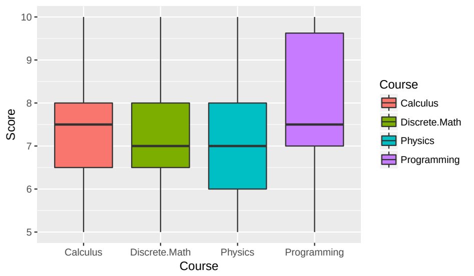
2019

Example (Student grades)



Student	Calculus	Discrete math	Physics	Programming
1	8.0	7.0	6.5	7
2	6.5	6.0	6.0	7
3	8.0	8.0	6.5	10
4	7.5	7.5	7.0	7
5	6.0	6.5	6.0	10
6	7.5	8.0	8.0	8
...				

100 students, 4 courses



Questions we are interested in:

Which course is the most difficult?

Is Programming the easiest?

Are the math courses the most similar?

Two important questions

Can we answer our questions with a model based on the Normal distribution (univariate or multivariate)? Should we use some other distribution instead?

Why use MVN distribution and not just use independent univariate Normal models (one for each course)?

Multivariate Normal (Gaussian) distribution

Density of a k-variate Normal distribution:

$$p(\mathbf{y}) = \frac{1}{(2\pi)^{\frac{k}{2}}|\Sigma|^{\frac{1}{2}}} \exp\left[-\frac{1}{2}(\mathbf{y}-\mu)^T\Sigma^{-1}(\mathbf{y}-\mu)\right],$$

where data $\mathbf{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_k \end{bmatrix}$, mean vector $\mu = \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_k \end{bmatrix}$, and covariance matrix

$$\Sigma = E[(\mathbf{y}-\mu)(\mathbf{y}-\mu)^T] = \begin{bmatrix} \sigma_1^2 & \sigma_{1,2} & \dots & \sigma_{1,k} \\ \sigma_{2,1} & \sigma_2^2 & \dots & \sigma_{2,k} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{k,1} & \sigma_{k,2} & \dots & \sigma_k^2 \end{bmatrix}.$$

Any restrictions regarding μ or Σ?

Bayesian MVN model

Model (likelihood), use reparametrization $\Lambda = \Sigma^{-1}$:

$y_i|\mu, \Lambda \sim MVN_k(\mu, \Lambda^{-1})$

Conjugate prior:

$\mu|\Lambda \sim MVN_k(\mu_0, (\kappa_0\Lambda)^{-1})$

$\Lambda \sim \text{Wishart}(\nu, S_0)$

Posterior:

$\mu|\boldsymbol{y}, \Lambda \sim MVN_k(\mu_n, (\kappa_n\Lambda)^{-1})$

$\Lambda|\boldsymbol{y} \sim \text{Wishart}(\nu_n, S_n^{-1})$

$\mu_n = \frac{\kappa_0\mu_0 + n\overline{y}}{\kappa_0 + n}$

$\kappa_n = \kappa_0 + n$

$S_n = S_0 + \sum_{i=1}^{n}(y_i - \overline{y})(y_i - \overline{y})^T + \frac{\kappa_0 n}{\kappa_0 + n}(\mu_0 - \overline{y})(\mu_0 - \overline{y})^T$

$\nu_n = \nu_0 + n$