A Gaussian Mixture Model (GMM) is a probabilistic model that assumes the data is generated from a mixture of several Gaussian distributions, each with its own parameters. Formally, a random vector $X \in \mathbb{R}^m$ is said to follow a GMM if its density function is a weighted sum of multivariate Gaussian densities:

shupeng.chen@mail.utoronto.ca

$$f(\mathbf{x}) = \sum_{k=1}^{K} \pi_k N_m(\mathbf{x}; \mu_k, \Sigma_k), \tag{3.7}$$

where $\pi_k \geq 0$ are the mixture weights such that $\boxed{\sum_{k=1}^{K} \pi_k = 1}$, and $N_m(\mathbf{x}; \mu_k, \Sigma_k)$ denotes the multivariate normal density with mean $\mu_k \in \mathbb{R}^m$ and covariance matrix $\Sigma_k \in \mathbb{S}^m$

Exercise 3.5.25. *Equation (3.7) provides the density of the Gaussian mixture model. Verify this function indeed integrates to 1.*

W.T.S $\int_{\mathbb{R}^m} f(x)\, d = 1$ where $f(x) = \sum_{k=1}^{K} \pi_k N_m(x; \mu_k, \Sigma_k)$

$$= \sum_{k=1}^{K} \pi_k \int N_m(x; \mu_k, \Sigma_k)\, dx$$

$$= 1 \cdot 1 = 1$$

$\int N_m(x; \mu_k, \Sigma_k)\, dx = 1$ (Since it's Multivariate Gaussian Density function)

$\sum_{k=1}^{K} \pi_k = 1$ (by def).

$$\int_{\mathbb{R}^m} N_m(x; \mu_k, \Sigma_k)\, dx$$

$$= \int_{\mathbb{R}^m} \frac{1}{(2\pi)^{\frac{m}{2}} |\Sigma|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(x-\mu)^T \Sigma^{-1} (x-\mu)\right) dx.$$

Let $y = x - \mu$

$$= \frac{1}{(2\pi)^{\frac{m}{2}} |\Sigma|^{\frac{1}{2}}} \int_{\mathbb{R}^m} \exp\left(-\frac{1}{2} y^T \Sigma^{-1} y\right) dy$$

Let $\Sigma = LL^T$ where L is lower triangular Matrix, Define $z = L^{-1} y$

$$= \frac{1}{(2\pi)^{\frac{m}{2}} |\Sigma|^{\frac{1}{2}}} \int_{\mathbb{R}^m} \exp\left(-\frac{1}{2} z^T z\right) |\Sigma|^{\frac{1}{2}} dz.$$

$$= \frac{1}{(2\pi)^{\frac{m}{2}}} \int_{R^m} \exp\left(-\frac{1}{2} z^T z\right) dz$$

$$= \frac{1}{(2\pi)^{\frac{m}{2}}} \left(\int_{-\infty}^{\infty} \exp\left(-\frac{1}{2} z_i^2\right) dz_i \right)^m$$

$$= \frac{1}{(2\pi)^{\frac{m}{2}}} \cdot (2\pi)^{\frac{m}{2}} = 1 \quad \blacksquare$$

Exercise 3.5.29. *Consider the Gaussian mixture model with a fixed number of components K. Suppose that a sample $\mathbf{x}_1, \ldots, \mathbf{x}_n$ is observed and take the corresponding log-likelihood function. We will show that this function is not bounded.*

1. *Consider the density $f(x; \mu, \sigma^2)$ of the uniform normal $N(\mu, \sigma^2)$. What happens to $f(x; \mu, \sigma^2)$ when $\sigma^2 \to 0$ for a fixed μ? Specifically, evaluate $f(x; \mu, \sigma^2)$ when $x = \mu$.*

2. *Now consider a GMM with $K = 2$ and $\pi_1 = 0.5, \pi_2 = 0.5$. Let $\mathbf{x}_1$ be one of the data points. Show that if one component, say $N_m(x; \boldsymbol{\mu}_1, \Sigma_1)$, collapses onto $\mathbf{x}_1$ (i.e., $\boldsymbol{\mu}_1 = \mathbf{x}_1, \Sigma_1 \to \mathbf{0}$), then the likelihood becomes arbitrarily large.*

1. $$f(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

when $x = \mu$

$$= \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^0.$$

$$= \infty \quad \text{as} \quad \sigma^2 \to 0.$$

2. $$l(\theta) = \sum_{i=1}^{n} \log\left(\sum_{k=1}^{2} \pi_k N_m(x_i; \mu_k; \Sigma_k)\right).$$

Suppose $N_m(x; \mu_1, \Sigma_1)$ collapase into x_1, that is $\mu_1 = x_1$, $\Sigma_1 \to 0$

then $$N_m(x_1; \mu_1, \Sigma_1) = \frac{1}{(2\pi)^{\frac{m}{2}} |\Sigma_1|^{\frac{1}{2}}}$$

As $\Sigma_1 \to 0$ $|\Sigma_1| \to 0$ and $\frac{1}{|\Sigma_1|^{\frac{1}{2}}} \to \infty$

So $\log(\pi_1 N_m(x_1; \mu_1; \Sigma_1) + \pi_2 N_m(x_1; \mu_2; \Sigma_2)) \approx \log(0.5 \cdot \infty) = \infty$.

Principal Component Analysis

The primary challenge in multivariate statistics is managing a large number of variables, potentially in the millions. One approach is to model these variables directly using high-dimensional statistical techniques. Alternatively, dimensionality reduction methods can be applied to derive a smaller set of variables that capture the most significant relationships in the data. These derived variables often serve as effective substitutes for the original data.

Exercise 4.8.2. *Suppose you apply PCA to a dataset and find that the first two principal components explain 95% of the variance. What does this tell you about the structure of the data?*

Exercise 4.8.3. *Let $X \sim N_m(0, \Sigma)$. Show that PCA transformations preserve normality, i.e., the principal components also follow a normal distribution. Show that in this case the principal components are independent.*

4.8.2 ① Redundant Features / highly correlated.

② Strong linear Relationships.

4.8.3

① Given $X \sim N_m(0, \Sigma)$

$\Sigma = U \Lambda U^T$ (by eigendecomposition)

U is orthogonal Matrix whose coluns are the eigenvectors of Σ

Λ is diagonal Matrix with eigenvalues of Σ on digonal.

Priciple componentes : $Z = U^T X$

$\because X \sim N_m(0, \Sigma)$ we have $AX \sim N_m(0, A\Sigma A^T)$

and have $A = U^T$

So $Z = U^T X \sim N_m(0, U^T \Sigma U)$

$U^T \Sigma U = U^T (U \Lambda U^T) U = \Lambda$ by eigendecoposition

So $Z \sim N_m(0, \Lambda)$ ∎

② i Since $Z \sim N_m(0, \Lambda)$

we have components of Z are uncorrelated | $Cov(Z_i, Z_j) = 0$

ii) Since Z is also Multivarice Normal

by i, ii we have Z is independent ∎

Note : $f(x) = \frac{1}{(2\pi)^{\frac{m}{2}} \prod_{i=1}^{m} \sigma_i} \exp\left(-\frac{1}{2}\sum_{i=1}^{m}\left(\frac{x_i - \mu_i}{\sigma_i}\right)^2\right) = \prod_{i=1}^{m} \frac{1}{\sqrt{2\pi}\sigma_i} \exp\left(-\frac{1}{2}\left(\frac{x_i - \mu_i}{\sigma_i}\right)^2\right)$

There is, $f_{X,Y}(x,y) = f_X(x) f_Y(y)$ def of Independence.

Exercise 4.8.6. *Show that the principal components are uncorrelated by computing the covariance matrix of the transformed variables.*

Exercise 4.8.19. *Show that the matrix AB with $A \in \mathbb{R}^{k\times l}$ and $B \in \mathbb{R}^{l\times m}$ has rank $\leq l$.*

4.8.6 Suppose $X \sim N(\mu, \Sigma)$, U's columns are eigenvector of Σ

$Z = U^T(X-\mu)$

$Cov(Z) = E[(Z - E(Z))(Z - E(Z))^T]$.

$= E[ZZ^T]$

Since $Z = U^T(X-\mu)$

$E(X) = \mu$

$E(Z)$

$= E[U^T(X-\mu)]$

$= U^T E(X-\mu) = U^T 0 = 0$

Sub $Z = U^T(X-\mu)$

$Cov(Z) = E[(U^T(X-\mu))(U^T(X-\mu))^T]$

$= E[U^T(X-\mu)(X-\mu)^T U]$

$= U^T E[(X-\mu)(X-\mu)^T] U$

$= U^T \Sigma U$

$\because \Sigma = U \Lambda U^T$

$\therefore Cov(Z) = U^T U \Lambda U^T U = \Lambda$

So Covariance Matrix of Z is diagonal Matrix, There is $Cov(Z_i, Z_j) = 0$ 証

4.8.19 given $A \in R^{k \times l}$, $B \in R^{l \times m}$

Consider $B = [b_1, b_2 \cdots b_m]$ where b_i are columns of B.

$AB = A[b_1, b_2 \cdots b_m] = [Ab_1, \cdots Ab_m]$

each column of AB is a linear combination of the colum of A

Column space of $AB \subseteq$ column space of A

So $\text{Rank}(AB) \leq \text{Rank}(A)$

Recall that, Dimension of column space of a Matrix = Rank (Matrix)

$\Rightarrow \quad \text{Rank}(AB) \leq \text{Ran}(A) \leq l$ ∎

R Code pulled from https://www.statmethods.net/advgraphs/ggplot2.html

Install ggplot2 if not already installed

install.packages("ggplot2")

Load ggplot2 for visualization

library(ggplot2)

In [4]:
```r
# Load the iris dataset
data(iris)

# View the first few rows of the dataset
head(iris)
```

A data.frame: 6 × 5

	Sepal.Length	Sepal.Width	Petal.Length	Petal.Width	Species
	<dbl>	<dbl>	<dbl>	<dbl>	<fct>
1	5.1	3.5	1.4	0.2	setosa
2	4.9	3.0	1.4	0.2	setosa
3	4.7	3.2	1.3	0.2	setosa
4	4.6	3.1	1.5	0.2	setosa
5	5.0	3.6	1.4	0.2	setosa
6	5.4	3.9	1.7	0.4	setosa

In [5]:
```r
# Perform PCA
pca_result <- prcomp(iris[, 1:4], center = TRUE, scale. = TRUE)

# Summary of PCA
summary(pca_result)
```

```
Importance of components:
                          PC1    PC2     PC3     PC4
Standard deviation     1.7084 0.9560 0.38309 0.14393
Proportion of Variance 0.7296 0.2285 0.03669 0.00518
Cumulative Proportion  0.7296 0.9581 0.99482 1.00000
```

In [6]:
```r
# Standard deviations of the principal components
pca_result$sdev

# Rotation (loadings)
pca_result$rotation

# Principal component scores
pca_result$x
```

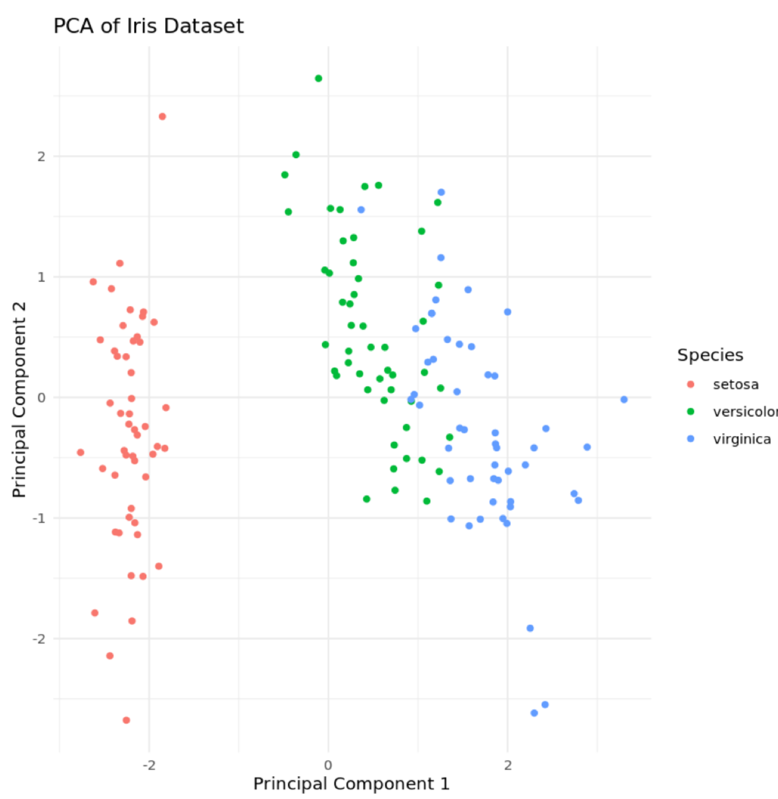
1.70836114932762 · 0.956049408486857 · 0.38308860015839 · 0.143926496617611

A matrix: 4 × 4 of type dbl

	PC1	PC2	PC3	PC4
Sepal.Length	0.5210659	-0.37741762	0.7195664	0.2612863
Sepal.Width	-0.2693474	-0.92329566	-0.2443818	-0.1235096
Petal.Length	0.5804131	-0.02449161	-0.1421264	-0.8014492
Petal.Width	0.5648565	-0.06694199	-0.6342727	0.5235971

A matrix: 150 × 4 of type dbl

PC1	PC2	PC3	PC4
-2.257141	-0.47842383	0.127279624	0.024087508
-2.074013	0.67188269	0.233825517	0.102662845
-2.356335	0.34076642	-0.044053900	0.028282305
-2.291707	0.59539986	-0.090985297	-0.065735340
-2.381863	-0.64467566	-0.015685647	-0.035802870
-2.068701	-1.48420530	-0.026878250	0.006586116
-2.435868	-0.04748512	-0.334350297	-0.036652767
-2.225392	-0.22240300	0.088399352	-0.024529919
-2.326845	1.11160370	-0.144592465	-0.026769540
-2.177035	0.46744757	0.252918268	-0.039766068
-2.159077	-1.04020587	0.267784001	0.016675503
-2.318364	-0.13263400	-0.093446191	-0.133037725
-2.211044	0.72624318	0.230140246	0.002416941
-2.624309	0.95829635	-0.180192423	-0.019151375
-2.191399	-1.85384655	0.471322025	0.194081578
-2.254661	-2.67731523	-0.030424684	0.050365010
-2.200217	-1.47865573	0.005326251	0.188186988
-2.183036	-0.48720613	0.044067686	0.092779618
-1.892233	-1.40032757	0.373093377	0.060891973
-2.335545	-1.12408360	-0.132187626	-0.037630354
-1.907931	-0.40749058	0.419885937	0.010884821
-2.199644	-0.92103587	-0.159331502	0.059398340
-2.765081	-0.45681330	-0.331069982	0.019582826
-1.812597	-0.08527285	-0.034373442	0.150636353
-2.219727	-0.13679618	-0.117599566	-0.269238379
-1.945329	0.62352971	0.304620475	0.043416203
-2.044303	-0.24135499	-0.086075649	0.067454082
-2.161336	-0.52538942	0.206125707	0.010241084
-2.132420	-0.31217200	0.270244895	0.083977887
-2.257698	0.33660425	-0.068207276	-0.107918349
⋮	⋮	⋮	⋮
2.0309126	-0.90742744	-0.234015510	0.167390481
0.9747153	0.56985526	-0.825362161	0.027662914
2.8879765	-0.41225995	0.854558973	-0.126911337
1.3287806	0.48020250	0.005410239	0.139491837
1.6950553	-1.01053648	-0.297454114	-0.061437911
1.9478014	-1.00441272	0.418582432	-0.217609339
1.1711801	0.31533806	-0.129503907	0.125001677
1.0175417	-0.06413118	-0.336588365	-0.008625505
1.7823788	0.18673563	-0.269754304	0.030983849
1.8574250	-0.56041329	0.713244682	-0.207519953
2.4278203	-0.25841871	0.725386035	-0.017863520
2.2972318	-2.61755442	0.491826144	-0.210968943
1.8564838	0.17795333	-0.352966242	0.099675959
1.1104277	0.29194458	0.182875741	-0.185721512
1.1984584	0.80860636	0.164173760	-0.487849130
2.7894256	-0.85394254	0.541093785	0.294893130
1.5709929	-1.06501321	-0.942695700	0.035486875
1.3417970	-0.42102015	-0.180271551	-0.214702016
0.9217370	-0.01716559	-0.415434449	0.005220919
1.8458612	-0.67387065	0.012629804	0.194543500
2.0080832	-0.61183593	-0.426902678	0.246711805
1.8954342	-0.68727307	-0.129640697	0.468128374
1.1540156	0.69653640	-0.528389994	-0.040385459
2.0337450	-0.86462403	-0.337014969	0.045036251
1.9914755	-1.04566567	-0.630301866	0.213330527
1.8642579	-0.38567404	-0.255418178	0.387957152
1.5593565	0.89369285	0.026283300	0.219456899
1.5160915	-0.26817075	-0.179576781	0.118773236
1.3682042	-1.00787793	-0.930278721	0.026041407
0.9574485	0.02425043	-0.526485033	-0.162533529



In [7]:
```r
# Create a data frame with the first two principal components and the species information
pca_data <- data.frame(PC1 = pca_result$x[, 1], PC2 = pca_result$x[, 2], Species = iris$Species)

# Plot the first two principal components
ggplot(pca_data, aes(x = PC1, y = PC2, color = Species)) +
  geom_point() +
  labs(title = "PCA of Iris Dataset", x = "Principal Component 1", y = "Principal Component 2") +
  theme_minimal()
```