Mixture Discriminant Analysis: A supervised GMM model

CWRU CSDS491 Spring 2024
Instructor: Dr. Michael Lewicki
Quan Le

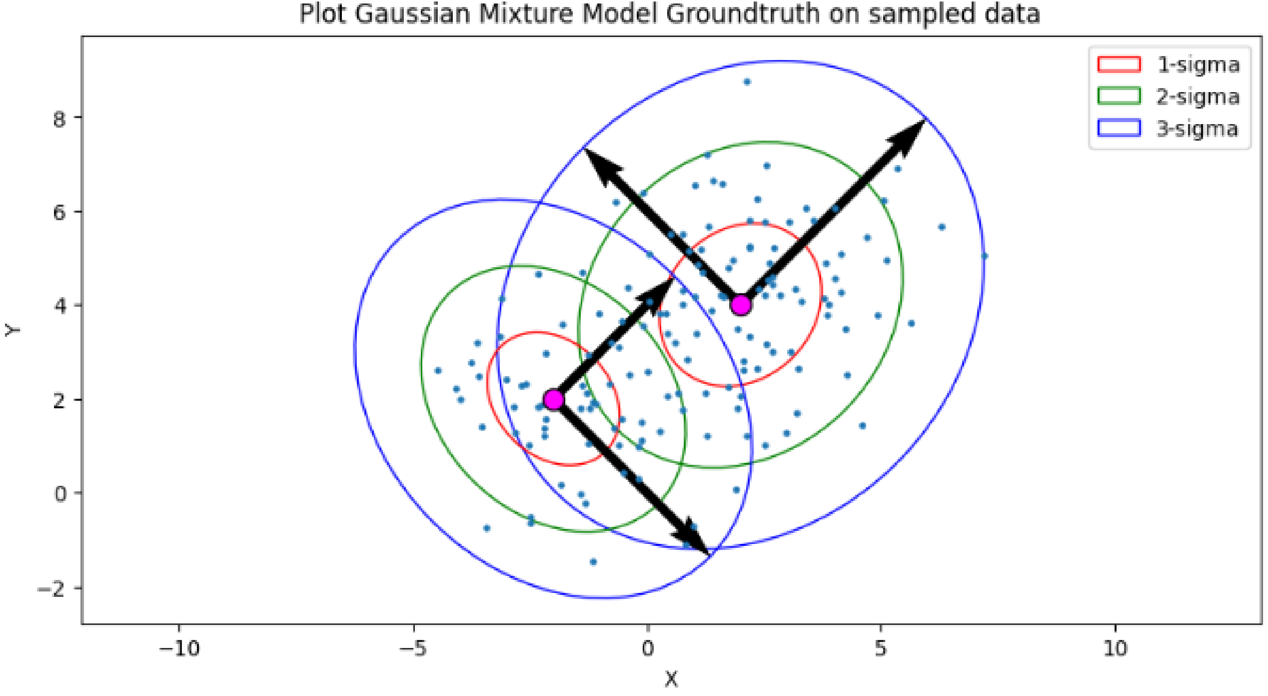
Content

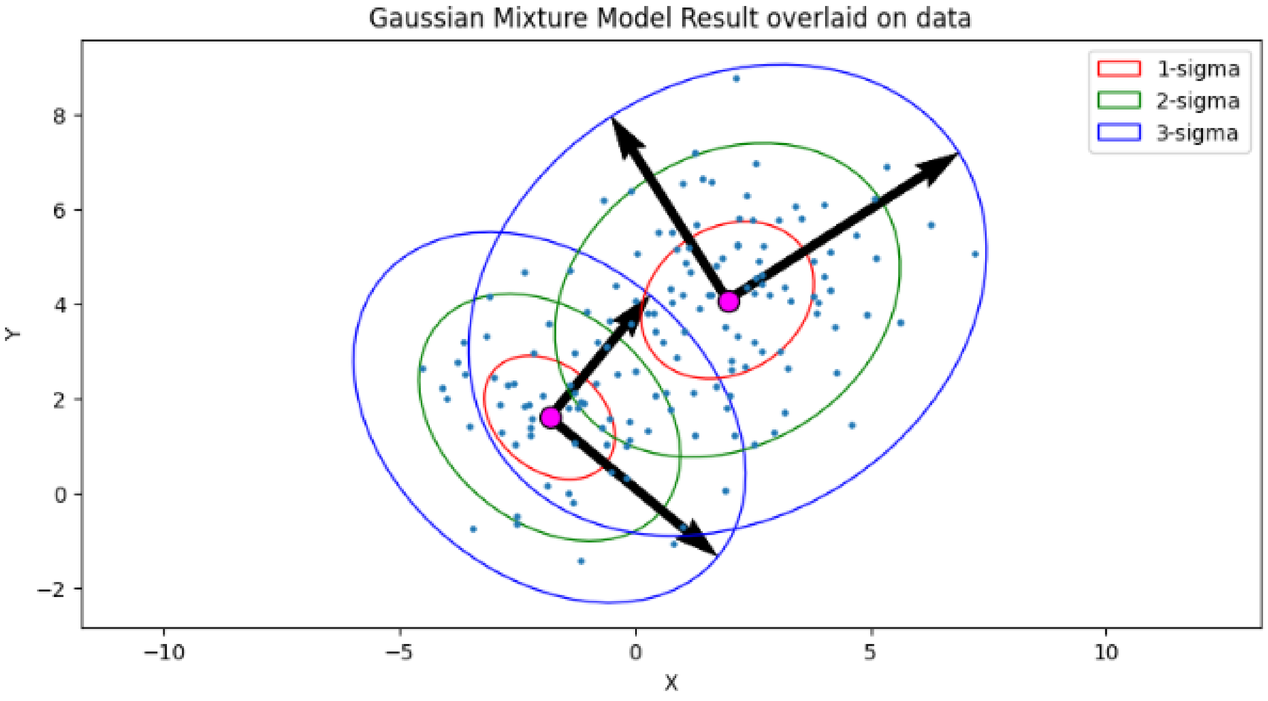
- Inspiration
- Collapsing the two worlds
- Mixture Discriminant Analysis
- Derivation
- Implementation, Results, and Evaluation

Inspiration: From cluster to class?

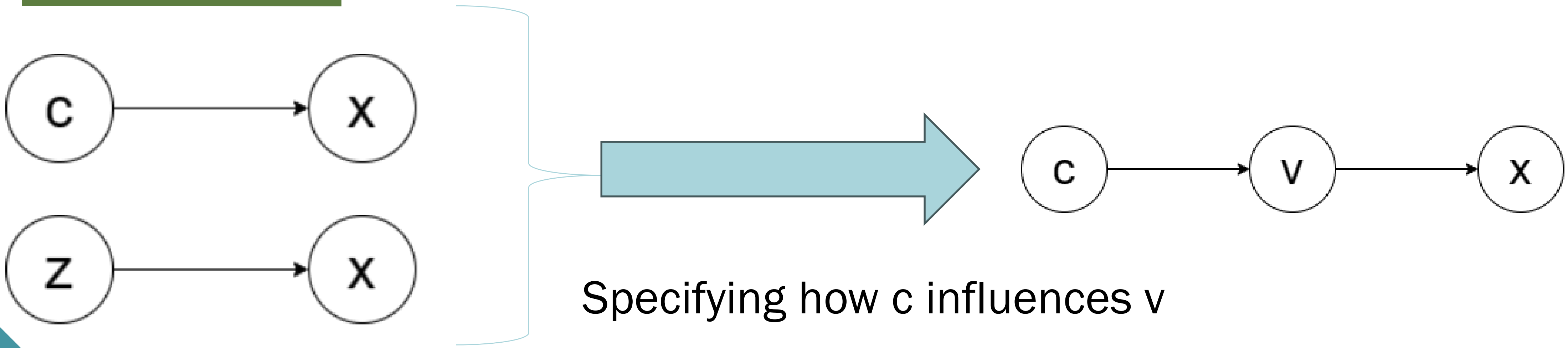
Gaussian Mixture Model is traditionally unsupervised

In a classification setting, what is the relationship then between the ground truth labels (class) and the notion of clusters?





Collapsing the two worlds



Specifying how c influences v

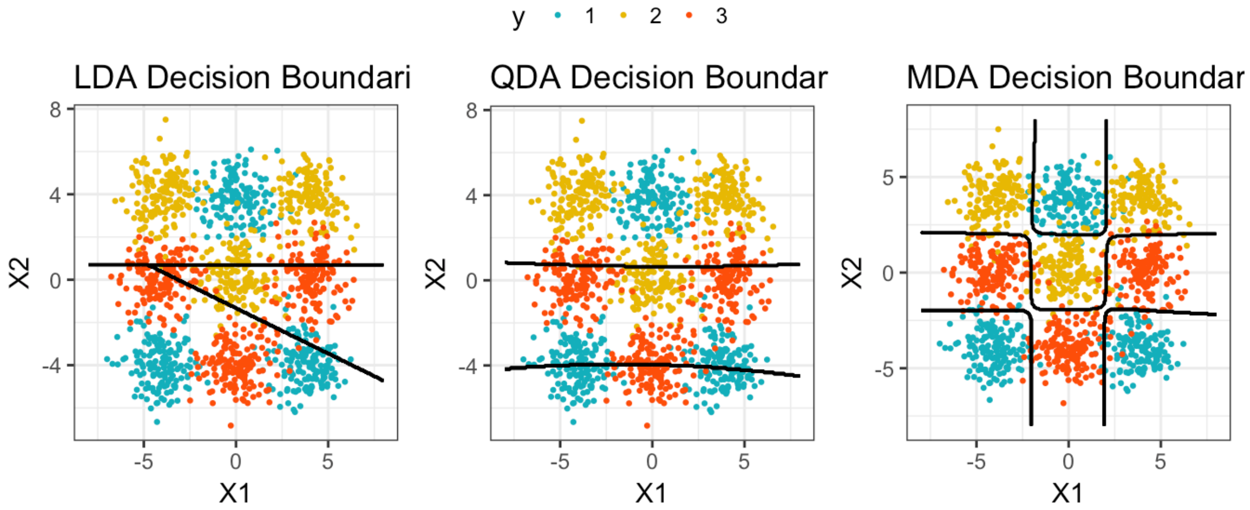
- One-to-one Mapping: Each class is modeled by 1 Gaussian Component (class decides Gaussian cluster)

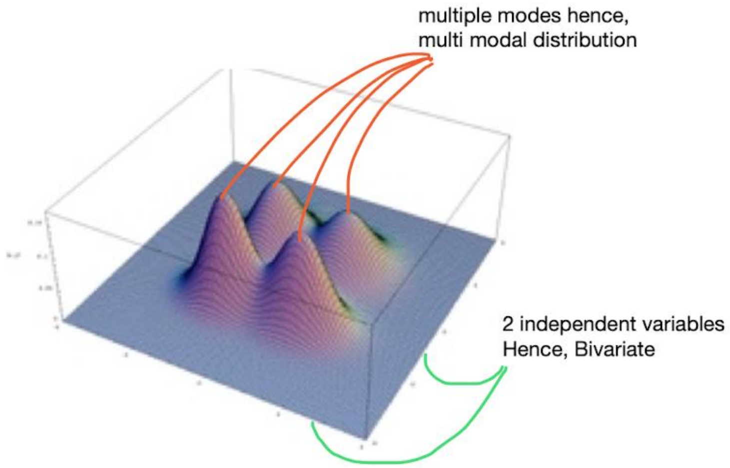
- Distribution: Each class is modeled by a mixture of Gaussian Components (cluster uncertainty)

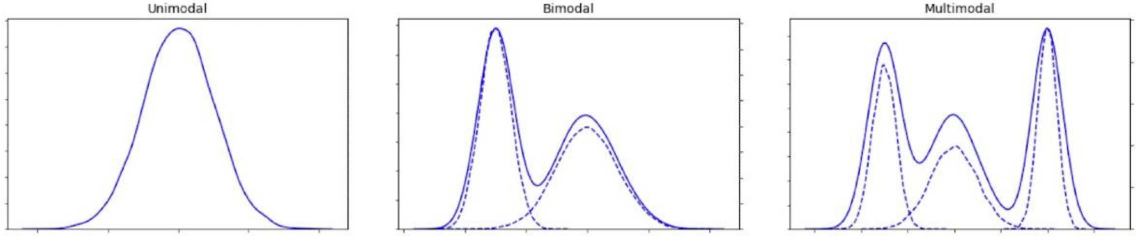
- Two different ways of modeling the class random variables produce 2 statistical learning methods

One-to-one: Quadratic Discriminant Analysis (QDA)

Distribution: Mixture Discriminant Analysis (MDA)







The model is only as good as the fit of its assumptions !!!

Derivation of MDA

- Derivation is the same process used to derive the unsupervised version, extended by adding a random variable class through Bayes' Rules
- Mixing coefficients are no more than probability
- Interpret everything as probability gives a coherent and intuitive understanding of terms

$$\mu_{ij} = \frac{1}{N_{ij}} \sum_{g_k=i} \gamma(z_{kij}) x_k$$

$$\Sigma_{ij} = \frac{1}{N_{ij}} \sum_{g_k=i} \gamma_{kij} (x_k - \mu_{ij})(x_k - \mu_{ij})^T$$

$$\pi_{ij} = \frac{N_{ij}}{N_i}$$

Hmm???!!! This seems familiar???

$$\text{With } \gamma(z_{kij}) = \frac{\pi_{ij} \mathcal{N}_{ij}(x_k)}{\sum_{r=1}^{R_i} \pi_{ir} \mathcal{N}_{ir}(x_k)} \text{ and } N_{ij} = \sum_{g_k=i} \gamma(z_{kij})$$

$$\boldsymbol{\mu}_k^{\text{new}} = \frac{1}{N_k} \sum_{n=1}^{N} \gamma(z_{nk}) \mathbf{x}_n \quad (9.24)$$

OH WAIT!!!

$$\boldsymbol{\Sigma}_k^{\text{new}} = \frac{1}{N_k} \sum_{n=1}^{N} \gamma(z_{nk}) \left(\mathbf{x}_n - \boldsymbol{\mu}_k^{\text{new}}\right) \left(\mathbf{x}_n - \boldsymbol{\mu}_k^{\text{new}}\right)^{\mathrm{T}} \quad (9.25)$$

$$\pi_k^{\text{new}} = \frac{N_k}{N} \quad (9.26)$$

where

$$N_k = \sum_{n=1}^{N} \gamma(z_{nk}).$$

3.5. Check for singularities (e.g. N

Dataset

Breast Cancer Wisconsin (Diagnostic) Dataset

- 569 instances, 30 real-value features
- 2 classes: 357 benign, 212 malignant

- Distinct subclass covariance implementation, which should allow the model to capture more complex decision boundary/distribution shapes
- Each class is modeled by the same number of Gaussian components (different from the derivation)
- K is everything – Model complexity + Model Assumption
- Not **Will it fit?** But **Will it overfit?**

Implementation

Result and Evaluation

- Here I will provide proof of overfitting as K is increased – To be finished in milestone 4

Future Direction

- Class-specific number of Gaussian Clusters
- Discriminative Dictionary Learning

References

- Hastie, T., & Tibshirani, R. (1996). Discriminant Analysis by Gaussian Mixtures. *Journal of the Royal Statistical Society. Series B (Methodological), 58*(1), 155–176. http://www.jstor.org/stable/2346171
- Ma, J., Gao, W. The supervised learning Gaussian mixture model. *J. of Comput. Sci. & Technol.* **13**, 471–474 (1998). https://doi.org/10.1007/BF02948506
- Dr. Michael Lewicki's lecture slides
- John Ramey: A Brief Look at Mixture Discriminant Analysis | R-bloggers