Analysis of Corrected Graph Convolutions

Robert Wang, Aseem Baranwal, Kimon Fontoulakis

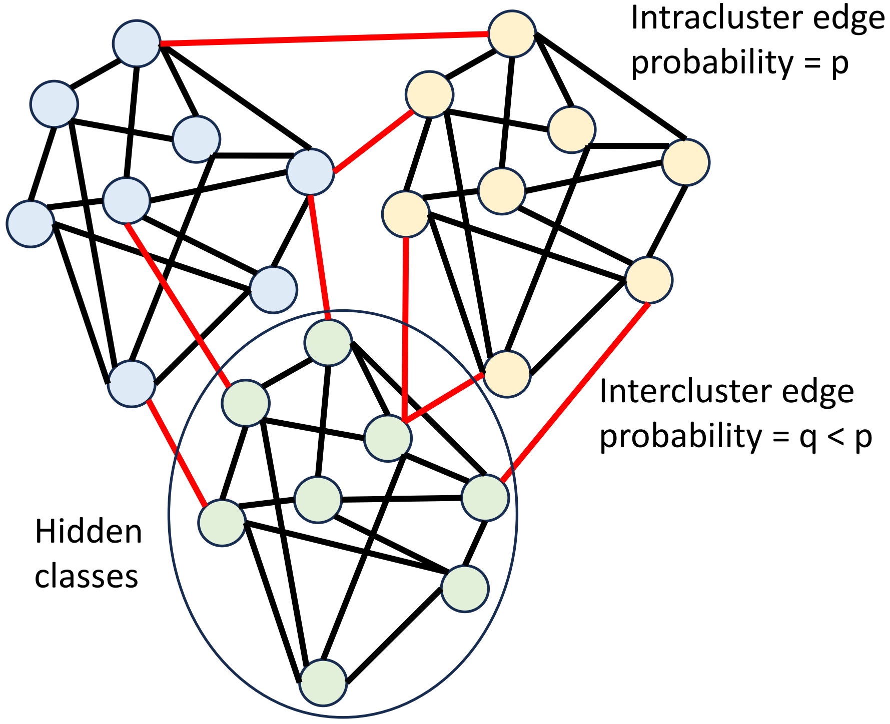
University of Waterloo

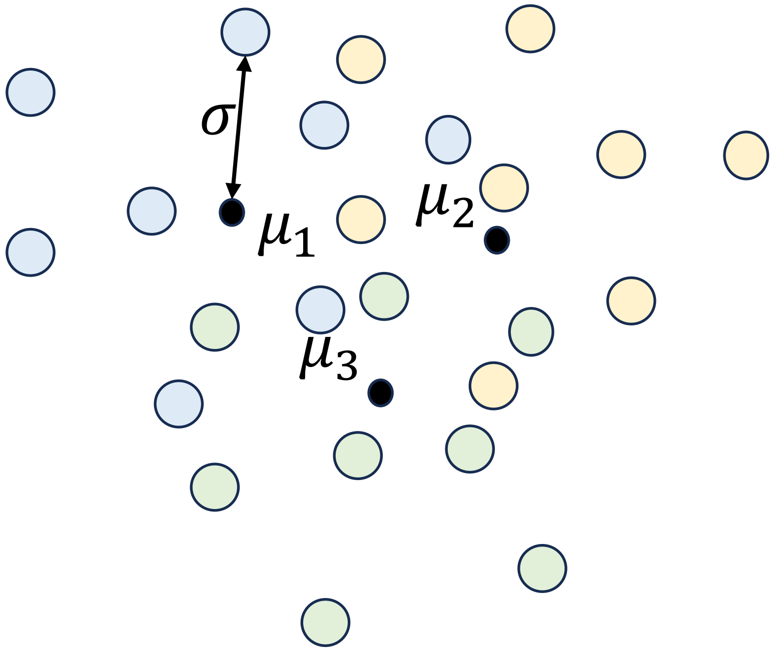
Graph Neural Networks (GNN)

- **Input**: A graph with an associated feature vector at each node
- **Assumption**: graph and the features are correlated and we wish to learn some signal from the input
- GNN takes the features as input to a neural network, and incorporates the graph into its architecture

Contextual Stochastic Block Model (CSBM)

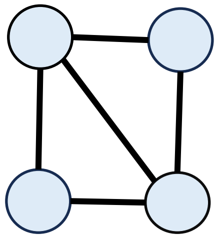
Graph ~ Stochastic Block Model + Features ~ Gaussian Mixture Model





Graph Convolution: $\underbrace{X}_{\text{Feature Matrix}} \mapsto \underbrace{M^k}_{\text{Convolution Matrix}} X$

Graph



Adjacency Matrix

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

Degree Matrix

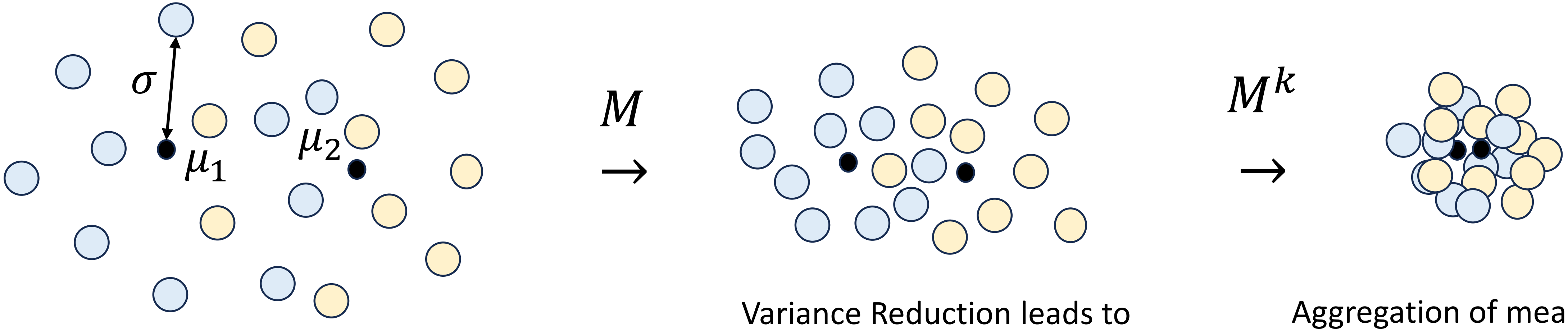
$$D = \begin{pmatrix} 3 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 2 \end{pmatrix}$$

Average degree $d = 2.5$

M matrix	Un-Normalized	Normalized
Standard	A	$D^{-\frac{1}{2}}AD^{-\frac{1}{2}}$
Corrected	$A - \underbrace{\frac{\mathrm{d}}{\mathrm{n}}11^\top}_{\text{“Expected” top eigenvector}}$	$D^{-\frac{1}{2}}AD^{-\frac{1}{2}} - \underbrace{\frac{1}{1^\top D1} D^{\frac{1}{2}}11^\top D^{\frac{1}{2}}}_{\text{Normalized top eigenvector}}$

Spectral Decomposition $M^k = \underbrace{\lambda_1^k v_1 v_1^\top}_{\text{Dominating signal}} + \underbrace{\sum_{i=2}^{n} \lambda_i^k v_i v_i^\top}_{\text{Corrected Convolution Matrix}}$

Oversmoothing in uncorrected Convolutions

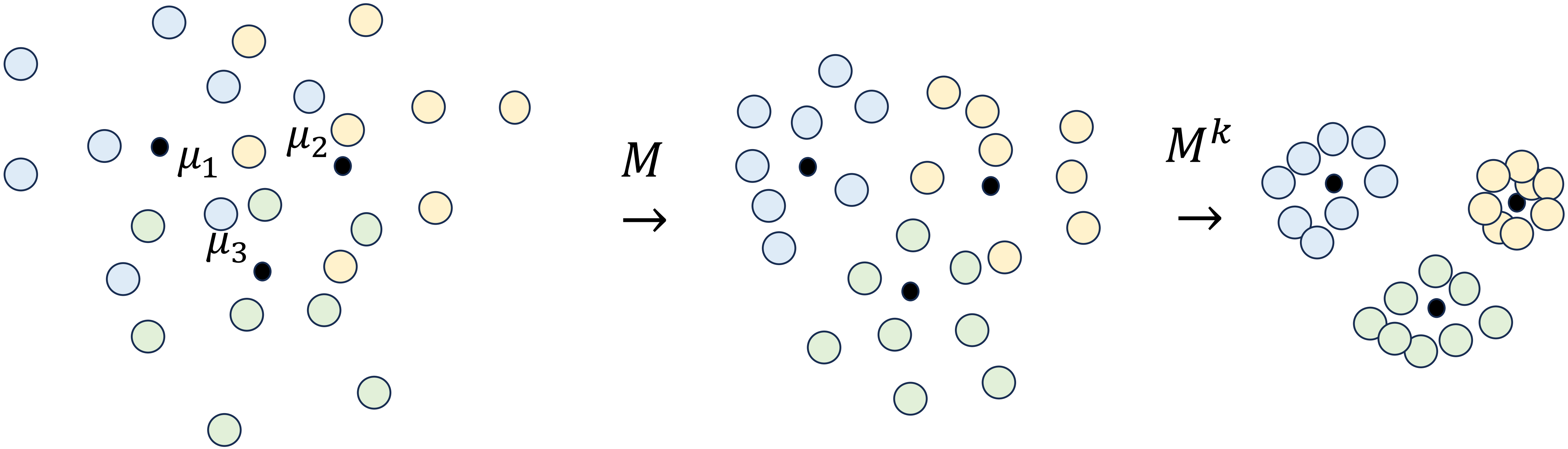


Variance Reduction leads to increased accuracy

Aggregation of means leads to decreased accuracy

Spectral Decomposition $M^k = \underbrace{\lambda_1^k v_1 v_1^\top}_{\text{Dominating Signal}} + \underbrace{\sum_{i=2}^{n} \lambda_i^k v_i v_i^\top}_{\text{Corrected Convolution Matrix}}$

Corrected Convolution



This work: Removing the top eigenvector leads to variance reduction without aggregation of the means

Classifiers

Linear (binary) Classifier: $X \mapsto M^k X \underbrace{w}_{} + \underbrace{b}_{}$

Trainable parameters

- Data is *Linearly Separable* if all entries in one class are positive and all entries in the other class is negative
- For multi-class data, can apply a linear classifier to each class

Non-linear Classifier: $softmax\left(||x_i^{(k)} - \underbrace{c_1}_{}||^2, ||x_i^{(k)} - \underbrace{c_2}_{}||^2, \ldots\right)_{i=1}^{n}$

Trainable parameters

Where $x_1^{(k)}, \ldots x_n^{(k)}$ are rows of the matrix $M^k X$

Binary Classification: Partial Recovery

- Parameters

Graph Signal: $\gamma = \frac{(p-q)\sqrt{np}}{p+q}$

Feature Separation: $\Delta = ||\mu_1 - \mu_2||$

Binary Classification: Partial Recovery

- Parameters

Graph Signal: $\gamma = \frac{(p-q)\sqrt{np}}{p+q}$

Feature Separation: $\Delta = ||\mu_1 - \mu_2||$

- Assumptions

$$p + q \geq \Omega(\frac{\log^2 \mathrm{n}}{\mathrm{n}}) \qquad \gamma \geq \Omega(1) \qquad \frac{\Delta}{\sigma} \geq \sqrt{\frac{\log n}{n}}$$

Binary Classification: Partial Recovery

- Parameters

Graph Signal: $\gamma = \frac{(p-q)\sqrt{np}}{p+q}$

Feature Separation: $\Delta = ||\mu_1 - \mu_2||$

- Assumptions

$$p + q \geq \Omega(\frac{\log^2 \mathrm{n}}{\mathrm{n}}) \qquad \gamma \geq \Omega(1) \qquad \frac{\Delta}{\sigma} \geq \sqrt{\frac{\log n}{n}}$$

- **Main Result**: There is a linear classifier using k corrected convolutions that, with high probability, has error rate at most

$$O\left(\frac{1}{\gamma^2} + \left(\frac{C}{\gamma^2}\right)^{2k} \cdot \frac{\sigma^2}{\Delta^2} \log n \right)$$

Binary Classification: Exact Recovery

Suppose our parameter satisfy the addition assumptions that:

$$p + q \geq \Omega(\frac{\log^3 \mathrm{n}}{\mathrm{n}}) \qquad \gamma \geq \Omega(k\sqrt{\log n}) \qquad \frac{\Delta}{\sigma} \geq \left(\frac{C}{\gamma}\right)^{2k} \sqrt{\log n}$$

Then the features are linearly separable after k convolutions with high probability

Two-Class Experiments

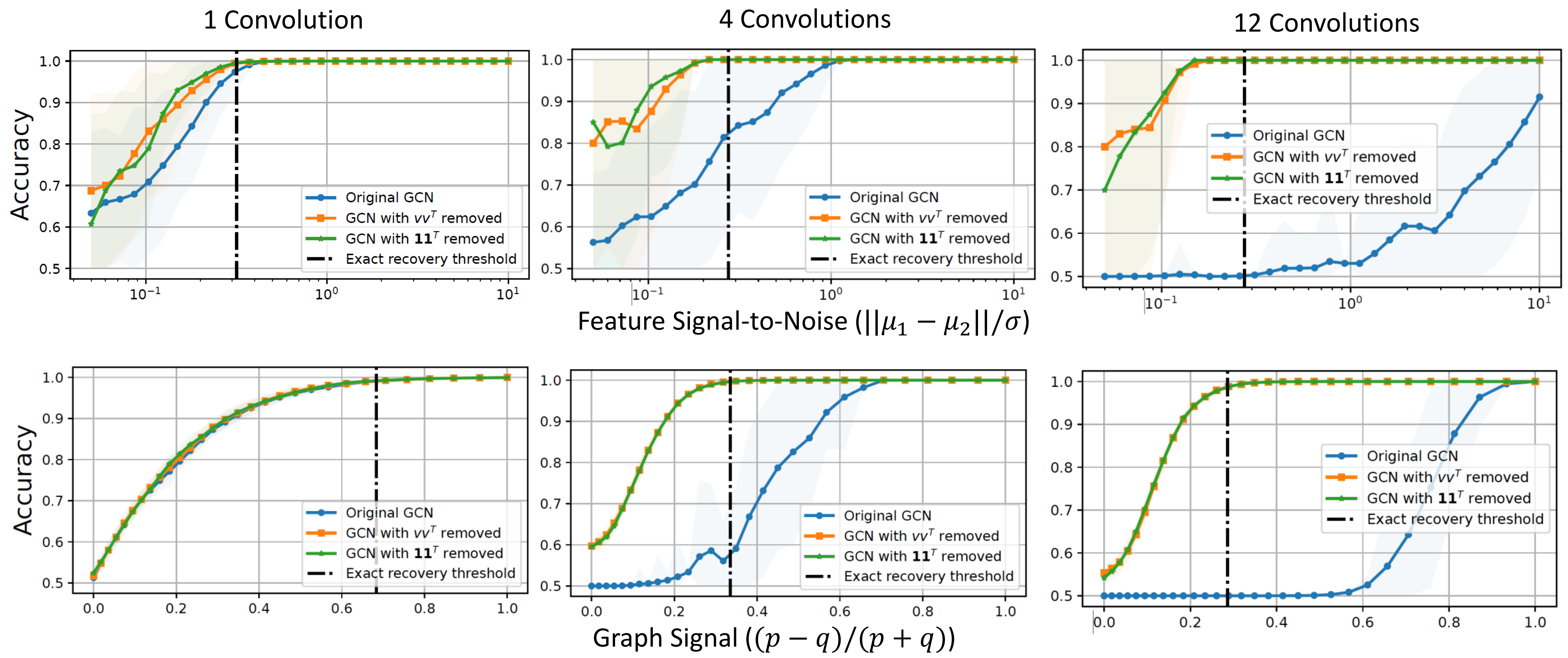


Figure: Synthetic experiments with $n = 2000$ and feature dimension 20 averaged over 50 trials. Green line is convolution with corrected un-normalized adjacency matrix. Orange line is convolution with normalized adjacency matrix, where v is its top eigenvector.

Multi-Class Classification: Partial Recovery

- Parameters: suppose we have L balanced classes with means $\mu_1, \dots \mu_L$

Graph Signal: $\gamma = \frac{(p-q)\sqrt{np}}{\sqrt{Lp(1-p)}+L\sqrt{q(1-q)}}$ **Feature Separation:** $\Delta = \min_{i,j} ||\mu_1 - \mu_2||$

- Assumptions

$$p + q \geq \Omega(\frac{\log^2 n}{n}) \qquad \gamma \geq \Omega(k) \qquad \frac{\Delta}{\sigma} \geq \sqrt{\frac{\log n}{n}}$$

- **Main Result**: There is a linear classifier using k corrected convolutions that, with high probability, has error rate at most

$$O\left(\frac{k^2}{\gamma^2} \cdot \frac{\sum_i ||\mu_i||^2}{L\Delta^2} + \left(\frac{L}{n} + \left(\frac{C}{\gamma^2}\right)^{2k} \right) \cdot \frac{\sigma^2}{\Delta^2} \log n \right)$$

Thanks for Watching