**Introduction**

**Motivation**
- The design of LDA heavily relies on the data covariance matrix which becomes ill conditioned in the large data regime.
- Most analysis focus on regularization techniques to overcome the high dimensionality effect on the estimation of the covariance matrix.
- Dimensionality reduction is an effective technique to get around high dimensionality but most analysis relies on bounds on the performance which might be loose in certain cases.
- Random projection is a common way to perform dimensionality reduction with some guarantees on the pairwise distances between data points (the Johnsonn-Lindenstrauss Lemma) but little can be told regarding the classification performance in terms of the statistics and the dimensions involved.

**Contributions**
- We consider LDA when data is randomly projected and arise from the multivariate Gaussian distribution.
- We investigate the classification performance for general random projection matrices satisfying some finite moments assumptions.
- We carry out the analysis when both the data dimension $d$ grow large simultaneously at the same rate, i.e. $d/p \to 0$.
- Under some mild assumptions controlling the data statistics, we show that the classification risk converges to a universal limit that describes in closed form fashion the performance in terms of the statistics and the dimensions involved.
- The obtained results permits to analytically quantify the performance loss due to projection which allows to carefully choose the reduced dimension in order to achieve a certain desirable performance.

**LDA with Random Projections**

**LDA**
- For a data point $x \in \mathbb{R}^p$, we say that $x \in C_i$ if $x \sim N(\mu_i, \Sigma_i)$.
- When the data is Gaussian, LDA is a Bayes classifier in the sense it maximizes $P[C_i|x]$ for $i \in \{0, 1\}$. The LDA score is

\[
W_{\text{LDA}}(x) = \left( x - \frac{\mu_0 + \mu_1}{2} \right)^\top \Sigma^{-1} (\mu_1 - \mu_0) + \log \frac{\pi_0}{\pi_1} + \frac{\pi_0}{\pi_1} \cdot 0.
\]

**Random Projections**

Random projection consists in the following operation.

\[
\mathbb{R}^p \to \mathbb{R}^d,
\]

\[
x \to W x
\]

**Johnsson-Lindenstrauss Lemma**

For a given $d$ data points $x_1, \ldots, x_d$ in $\mathbb{R}^p$, $\epsilon \in (0, 1)$ and $d > \frac{8 \log d}{\epsilon^2}$, there exists a linear map $f : \mathbb{R}^p \to \mathbb{R}^d$ such that

\[
|1 - \epsilon| |x_i - x_j|^2 \leq |f(x_i) - f(x_j)|^2 \leq (1 + \epsilon) |x_i - x_j|^2.
\]

What about the classification risk?

Conditioning on the projection matrix $W$, we have

\[
W_{\text{P-LDA}} = \Phi \left( -\frac{1}{2} \sqrt{\mu_1 - \mu_0} W (W C W)^{-1} W \mu_1 + \frac{(\mu_1 - \mu_0) \log \frac{\pi_0}{\pi_1} + \pi_0}{\sqrt{\mu_1 - \mu_0} W (W C W)^{-1} W \mu_1} \right).
\]

**Technical Assumptions**

**Assumption 1. (Growth rate)**

As $p \to \infty$, we assume the following
- Data scaling: $0 < \inf i \leq \limsup \frac{2}{p} \leq 1$.
- Mean scaling: \( \mu = \mu_0 - \mu_1, \limsup |\mu| < \infty. \)
- Covariance scaling: $\limsup |C| < \infty$.

**Assumption 2. (Projection matrix)**

We shall assume that the projection matrix $W$ writes as $W = \frac{1}{\sqrt{d}} Z$ where the entries $Z_{ij}$ ($1 \leq i \leq d, 1 \leq j \leq p$) of $Z$ are centered with unit variance and independent identically distributed random variables satisfying the following moment assumption.

There exists $\epsilon > 0$, such that $E|Z_{ij}|^{4+\epsilon} < \infty.$

**Main Results**

A Fundamental Result in RMT

Under Assumptions 1 and 2, it allows to construct a deterministic equivalent of $\left( \frac{1}{\sqrt{n}} Z^2 \right) + \frac{1}{\sqrt{p}} b$ denoted by $\mathcal{Q}(t) \in \mathbb{R}^{p \times p}$ in the sense that

\[
a \left( \frac{1}{\sqrt{n}} Z^2 \right) + \frac{1}{\sqrt{p}} b \sim \mathcal{Q}(t) \sim \mathcal{O}(t).
\]

for all deterministic $a$ and $b$ in $\mathbb{R}^p$ with uniformly bounded Euclidean norms and $t > 0$. $\mathcal{Q}(t)$ is a deterministic matrix given by $\mathcal{Q}(t) = \left( I_p + \frac{1}{p} \mathcal{C} \right)^{-1}$, where $\delta(t)$ satisfies $\delta(t) = \frac{t}{2} b \mathcal{Q}(t)$.

**Proposition 1. (Asymptotic Performance)**

Under Assumptions 1 and 2, then for $i \in \{0, 1\}$ the conditional probability of misclassification in (4) converges in probability to a non trivial deterministic limit given by

\[
\mathcal{P}_{\text{LDA}}(i) = \Phi \left( -\frac{1}{2} \sqrt{\mu_1 - \mu_0} W (W C W)^{-1} W \mu_1 + \frac{(\mu_1 - \mu_0) \log \frac{\pi_0}{\pi_1} + \pi_0}{\sqrt{\mu_1 - \mu_0} W (W C W)^{-1} W \mu_1} \right) \to \mathbb{P}_{\text{P-LDA}}.
\]

**Special cases**

- Equal priors, i.e. $\pi_0 = \pi_1$.

\[
\mathcal{P}_{\text{LDA}}(i) = \Phi \left( -\frac{1}{2} \sqrt{\mu_1 - \mu_0} W (W C W)^{-1} W \mu_1 + \frac{(\mu_1 - \mu_0) \log \frac{\pi_0}{\pi_1} + \pi_0}{\sqrt{\mu_1 - \mu_0} W (W C W)^{-1} W \mu_1} \right) \to \mathbb{P}_{\text{P-LDA}}.
\]

- Equal priors and $C = I_p$.

\[
\mathcal{P}_{\text{LDA}}(i) = \Phi \left( -\frac{1}{2} \sqrt{\mu_1 - \mu_0} W (W C W)^{-1} W \mu_1 + \frac{(\mu_1 - \mu_0) \log \frac{\pi_0}{\pi_1} + \pi_0}{\sqrt{\mu_1 - \mu_0} W (W C W)^{-1} W \mu_1} \right) \to \mathbb{P}_{\text{P-LDA}}.
\]

As expected, there is a performance loss due to projection and it is analytically characterized by Proposition 1. Conversely, for a given desired performance $\tau$, we can determine the minimum $d$ such that $\mathcal{P}_{\text{LDA}} \leq \tau$.

**Experiments**

We consider Gaussian and Bernoulli projection matrices generated as follows.
- Gaussian: $W_{ij} \sim_{i.i.d} N(0, 1/\rho)$.
- Bernoulli: $W_{ij} = \left\lfloor \left( 1 - 2B_{ij} \right) \right\rfloor$ where $B_{ij} \sim_{i.i.d} \text{Bernoulli}(1/2)$.

**Synthetic data**

The data is generated using the Gaussian distribution with the following parameters.
- $\rho = 800$.
- $\mu_0 = 0_p$ and $\mu_1 = \frac{1}{p} 1_p$.
- $C = \left( 0.4 + \frac{1}{2} I_p \right) 1_p$.

**MNIST data**

$C_1$ is taken to be the digit 2 whereas $C_2$ is given by digit 3.

We obtain the data statistics by relying on sample estimates computed from the training data.

![Figure: Misclassification rate of randomly-projected LDA.](image)