Finite Sample Effects in Compressed Fisher's LDA

Robert J.Durrant Ata Kabán

School of Computer Science, University of Birmingham, Edgbaston, UK

Summary

We consider one of several finite sample effects that can occur when carrying out classification of data using Fisher's Linear Discriminant (FLD) on a random projection of the observations.
In forthcoming conference papers [1, 2] we bound the error of FLD when we work in a k-dimensional random projection of the (non-sparse) d-dimensional data, $k \ll d$. We prove that as long as the true and sample means belonging to a class lie on the same side of the decision boundary as one another, for good generalization performance of the projected classifier (on average over the choice of random projection matrix R) for an $m+1$ class problem it is sufficient to take the projection dimensionality to be $k \in \mathcal{O}(\log m)$.
However, what if the means agree in this way in the data space but not in the projected space? How likely is it that one of the means will be 'flipped' across the decision boundary by random projection? We answer this question by giving the exact probability that two vectors with angular separation $\theta \in [0, \pi/2]$ in the data space have angular separation $\theta_R > \pi/2$ in the randomly projected space.

Theorem - Flip probability

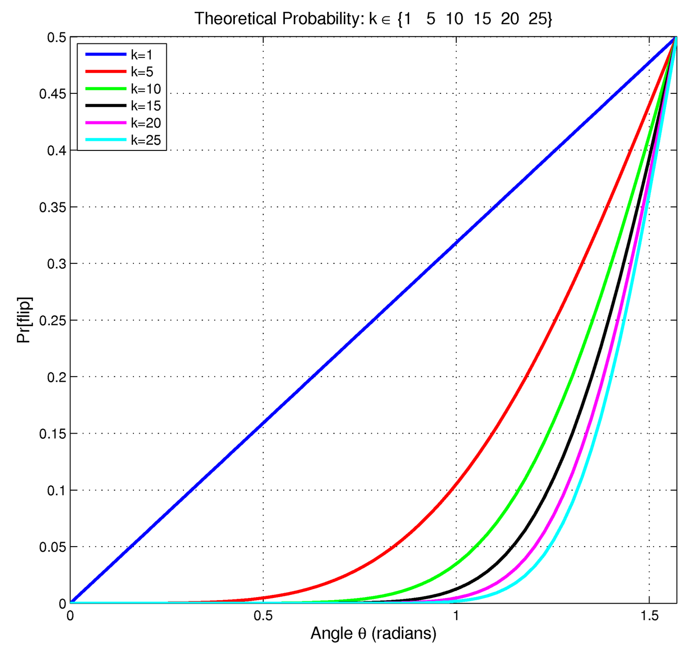
Let $\mathbf{n}, \mathbf{m} \in \mathbb{R}^d$ with angular separation $\theta \in [0, \pi/2]$.
Let $R \in \mathcal{M}_{k \times d}$ be a random projection matrix with entries $r_{ij} \overset{\text{iid}}{\sim} \mathcal{N}(0, 1/d)$ and let $R(\mathbf{n}), R(\mathbf{m}) \in \mathbb{R}^k$ be the projections of $\mathbf{n}, \mathbf{m}$ into $\mathbb{R}^k$ with angular separation θ_R.
Then the 'flip probability' $\Pr_R[\theta_R > \pi/2 | \theta] = \Pr_R[(R(\mathbf{n}))^T R(\mathbf{m}) < 0 | \mathbf{n}^T \mathbf{m} \geqslant 0]$ is given by:

$$\frac{\Gamma(k)}{(\Gamma(k/2))^2 2^{k-1}} \int_0^\theta \sin^{k-1}(\phi)\, d\phi \qquad (1)$$

Note that this probability is independent of d, the original data dimensionality!
Unfortunately no closed form exists for the integral in (1), but integrating by parts does give an exact expression for any given k.

Form of the flip probability (1)

The following graph shows the theoretical value of the flip probability for some different choices of projection dimension k:



Here we see that the flip probability drops very sharply as k increases (it is polynomial of order k in θ). For $k \geqslant 5$ the two vectors $\mathbf{m}$ and $\mathbf{n}$ must be separated by about 30° for the flip probability to be greater than machine precision. When these vectors are the true and sample mean of a Gaussian distribution, we see that the sample mean must be a very poor estimate of the true mean for flipping to occur if $k \geqslant 5$. This graph and the following corollary show that in practice it is unlikely we would ever have to evaluate the integral in (1) for large values of k.

Corollary to the theorem

Note that in each case plotted above, the flip probability for given k is dominated by every lower order flip probability. This can be shown to hold for all $\theta \in [0, \pi/2]$ and $k \in \mathbb{N}$, and hence we have the following corollary:
For $\theta \in [0, \pi/2]$ we have, for all k:

$$\frac{\Gamma(k)}{(\Gamma(k/2))^2 2^{k-1}} \int_0^\theta \sin^{k-1}(\phi)\, d\phi \geqslant \frac{\Gamma(k+1)}{(\Gamma((k+1)/2))^2 2^k} \int_0^\theta \sin^{k}(\phi)\, d\phi$$

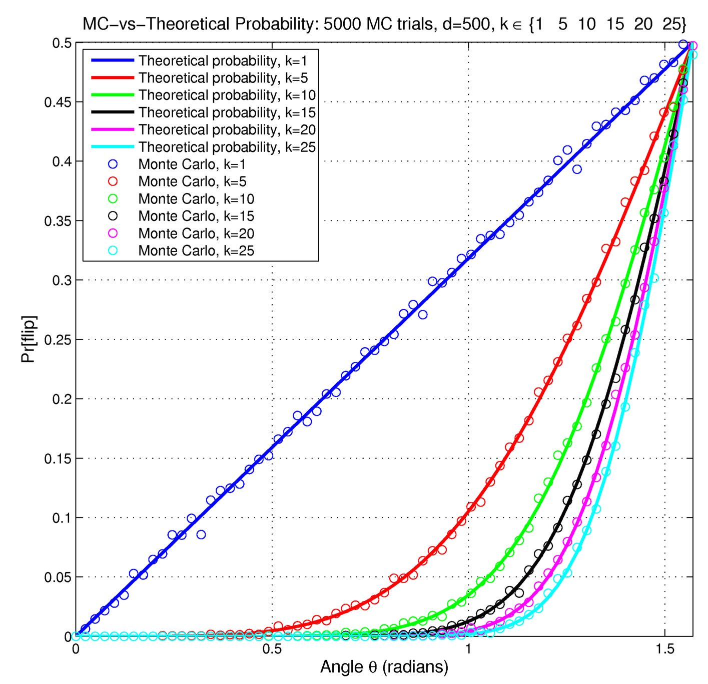
In particular, the flip probability for any k is bounded above by θ/π.

Monte Carlo trials

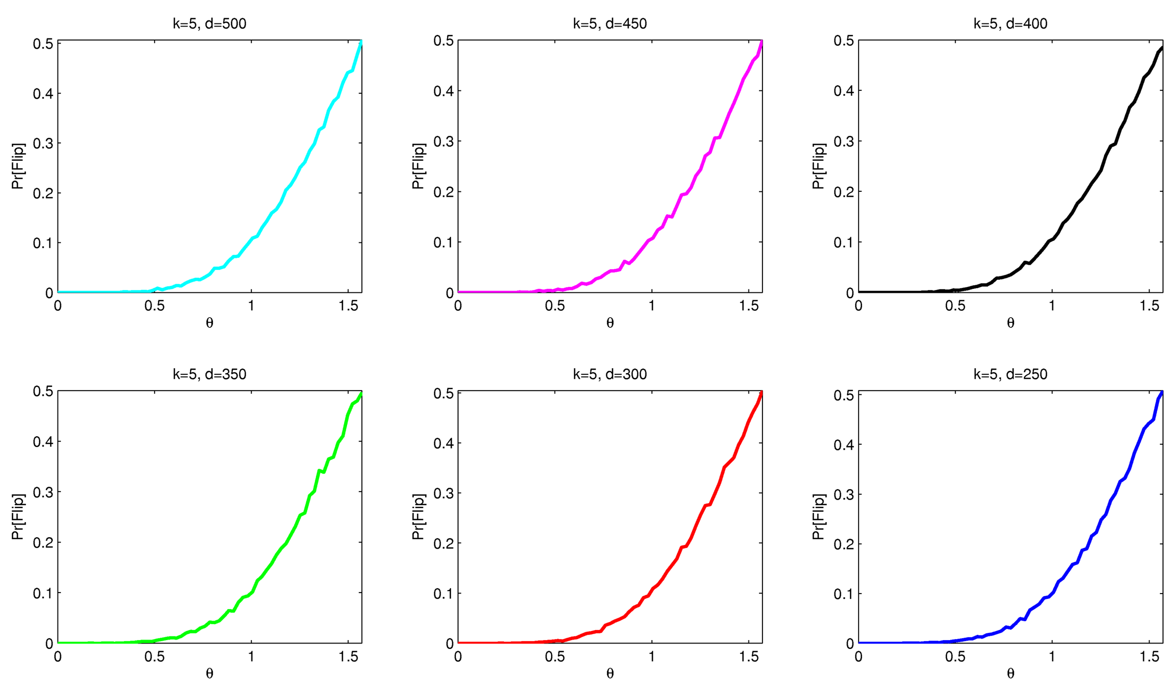
Our results seem quite counterintuitive, in particular, the fact that the flip probability is independent of the original dimensionality. To confirm our theoretical findings we ran Monte Carlo trials to estimate the flip probability as follows:

a. We let $d \in \{50, 100, \ldots, 500\}$, $k \in \{1, 5, 10, 15, 20, 25\}$ and $\theta \in \{0, \pi/128, \ldots, t \cdot \pi/128, \ldots, \pi/2\}$.
b. For each (d, θ) tuple we generated 2 randomly oriented d-dimensional θ-separated unit length vectors $\mathbf{m}, \mathbf{n}$.
c. For each (k, d, θ) tuple, we generated 5000 $k \times d$ random projection matrices R with which we randomly projected $\mathbf{m}$ and $\mathbf{n}$.
d. Finally we counted the number of times, N, that the dot product $(R(\mathbf{m}))^T R(\mathbf{n}) < 0$ and estimated the flip probability by $N/5000$.

The following plot shows the close match between our theoretical values and empirical estimates of the flip probabilities:



Empirical results confirm the flip probability is independent of d:

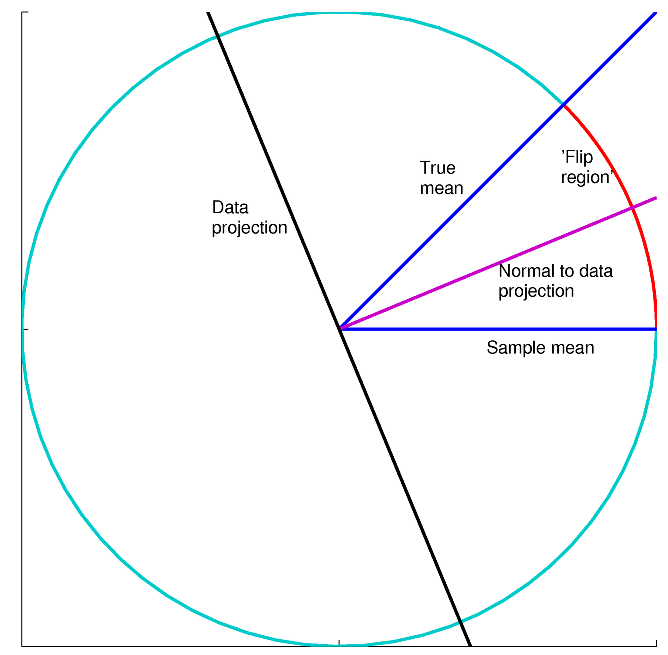


Geometric interpretation

Equation (1) can be recast as:

$$\frac{\Gamma(k)}{(\Gamma(k/2))^2 2^{k-1}} \int_0^\theta \sin^{k-1}(\phi)\, d\phi = \frac{\int_0^\theta \sin^{k-1}(\phi)\, d\phi}{\int_0^\pi \sin^{k-1}(\phi)\, d\phi}$$

which is the surface area in $\mathbb{R}^{k+1}$ of a hyperspherical cap divided by the surface area of the hypersphere. For $k = 1, d \geqslant 2$ this reduces to the following 2-dimensional situation:



If the orientation of the data projection is uniformly distributed on the unit circle, then we recover the $k = 1$ probability of θ/π.

Proof

The full proof will be available in technical report form at www.cs.bham.ac.uk/~durranrj in the near future.

References

- R.J. Durrant and A. Kabán.
Compressed Fisher Linear Discriminant Analysis: Classification of Randomly Projected Data.
In *Proceedings KDD 2010, to appear*, 2010.
- R.J.Durrant and A. Kabán.
A bound on the performance of LDA in randomly projected data spaces.
In *Proceedings ICPR 2010, to appear*, 2010.
- B. Eisenberg and R. Sullivan.
Random triangles in n dimensions.
American Mathematical Monthly, pages 308–318, 1996.

http://www.cs.bham.ac.uk/~durranrj

r.j.durrant@cs.bham.ac.uk