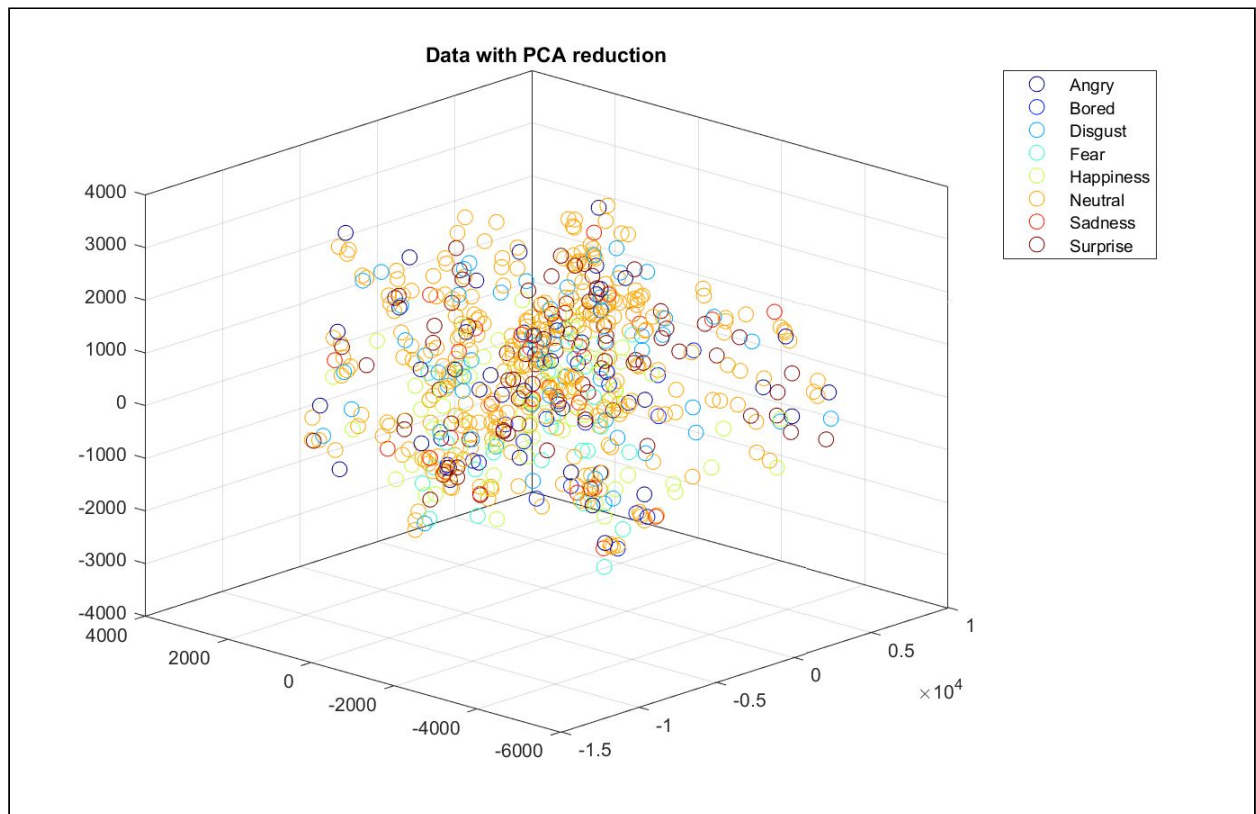


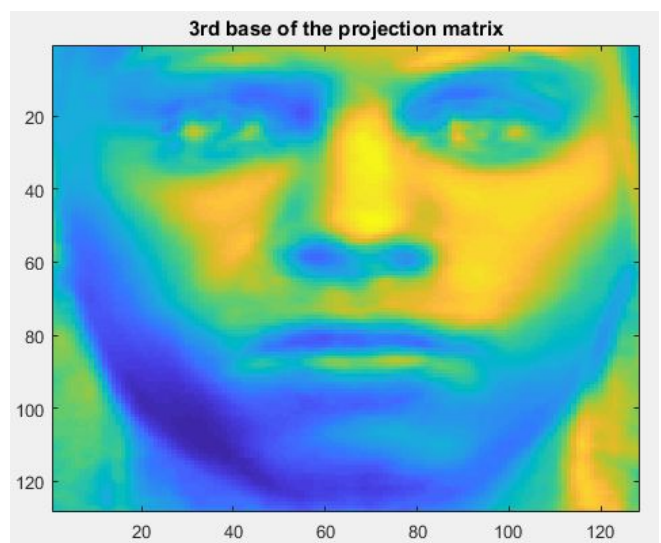
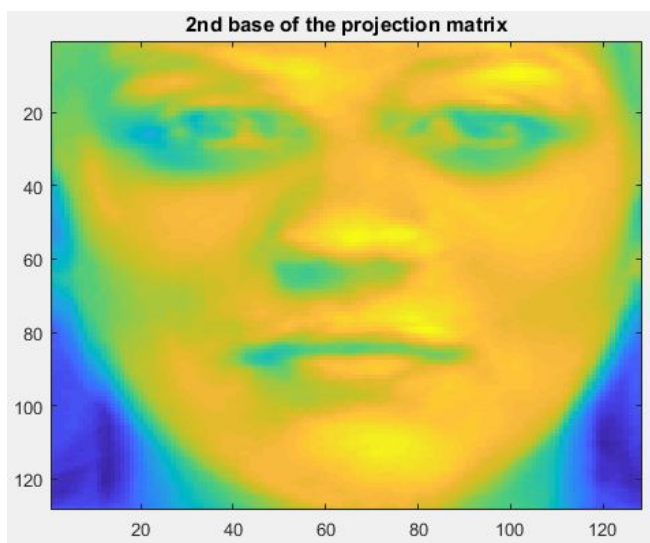
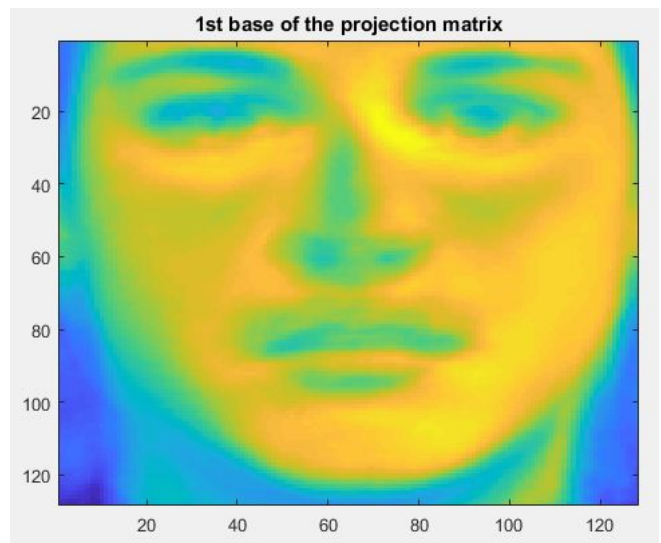
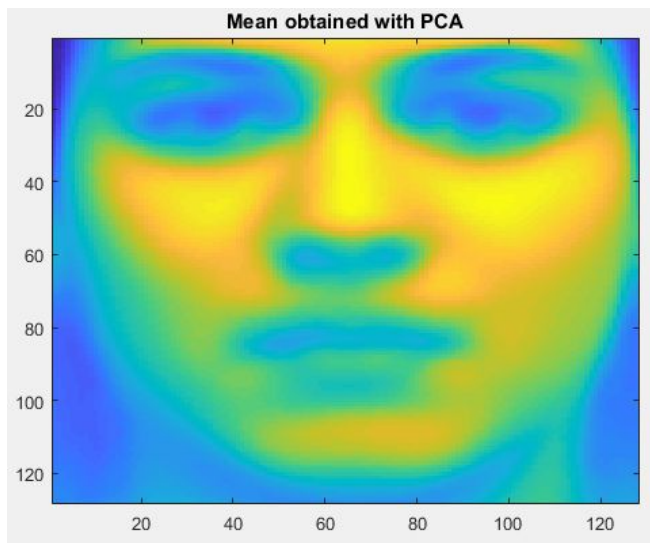
## PRÀCTICA 2: "Dimensionality reduction"

1. Per a poder visualitzar el següent plot hem fet una extracció dels descriptors:



Aquestes són les dades projectades després de reduir la dimensionalitat amb PCA.

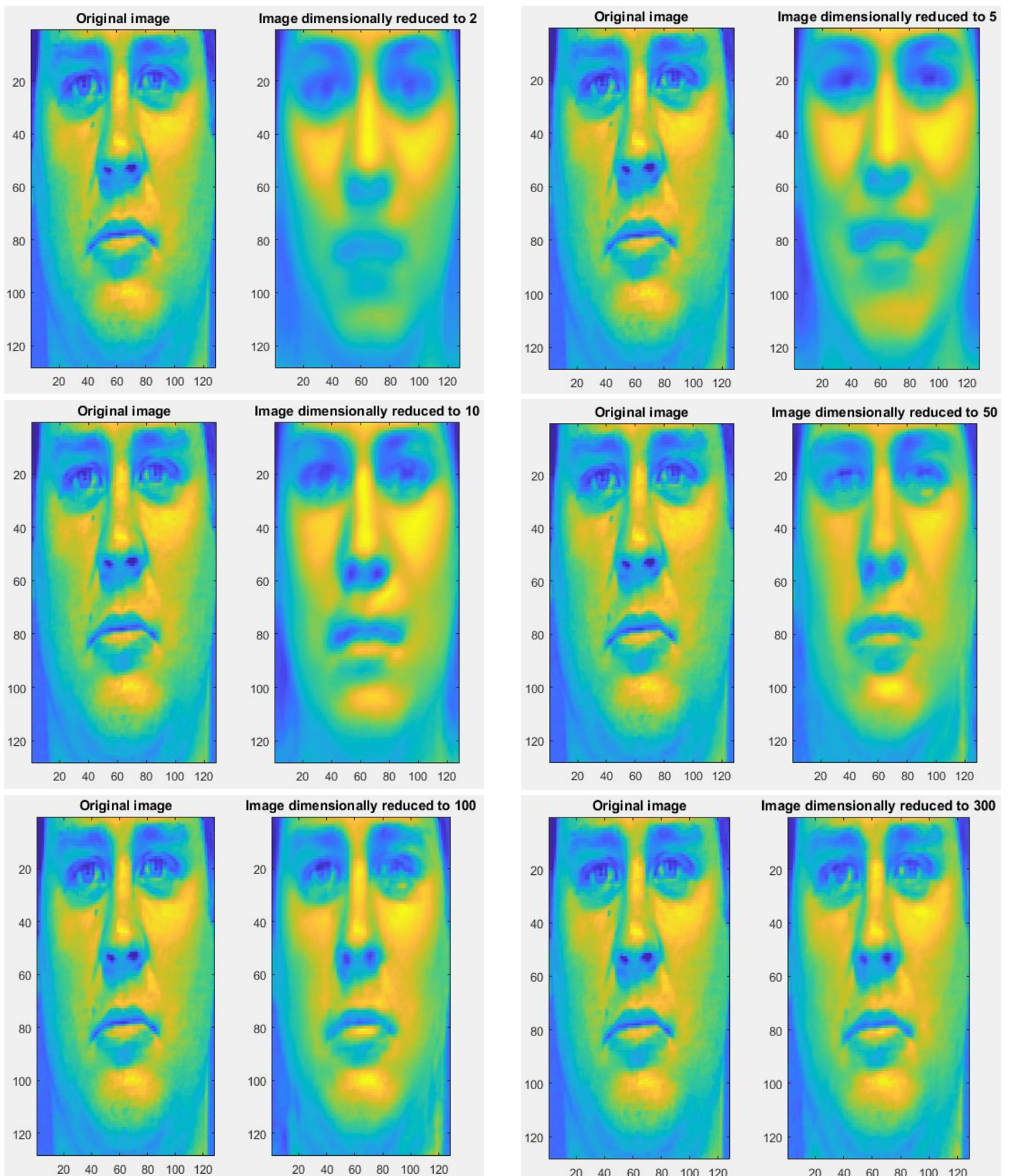
## 2.



La mitja que obtenim amb PCA és la mitja que fa de les imatges de la *database*, com el “resultat final” després de combinar-les totes. Com que hem reduït les dimensions de la base de dades en 3 dimensions, quan tinguem una imatge es podrà projectar amb aquestes 3 bases, que serviran com a eixos 3D per a la projecció de les dades; en aquest cas les dades són imatges i es projecten en unes bases que també ho són, com es pot veure en les fotografies anteriors.

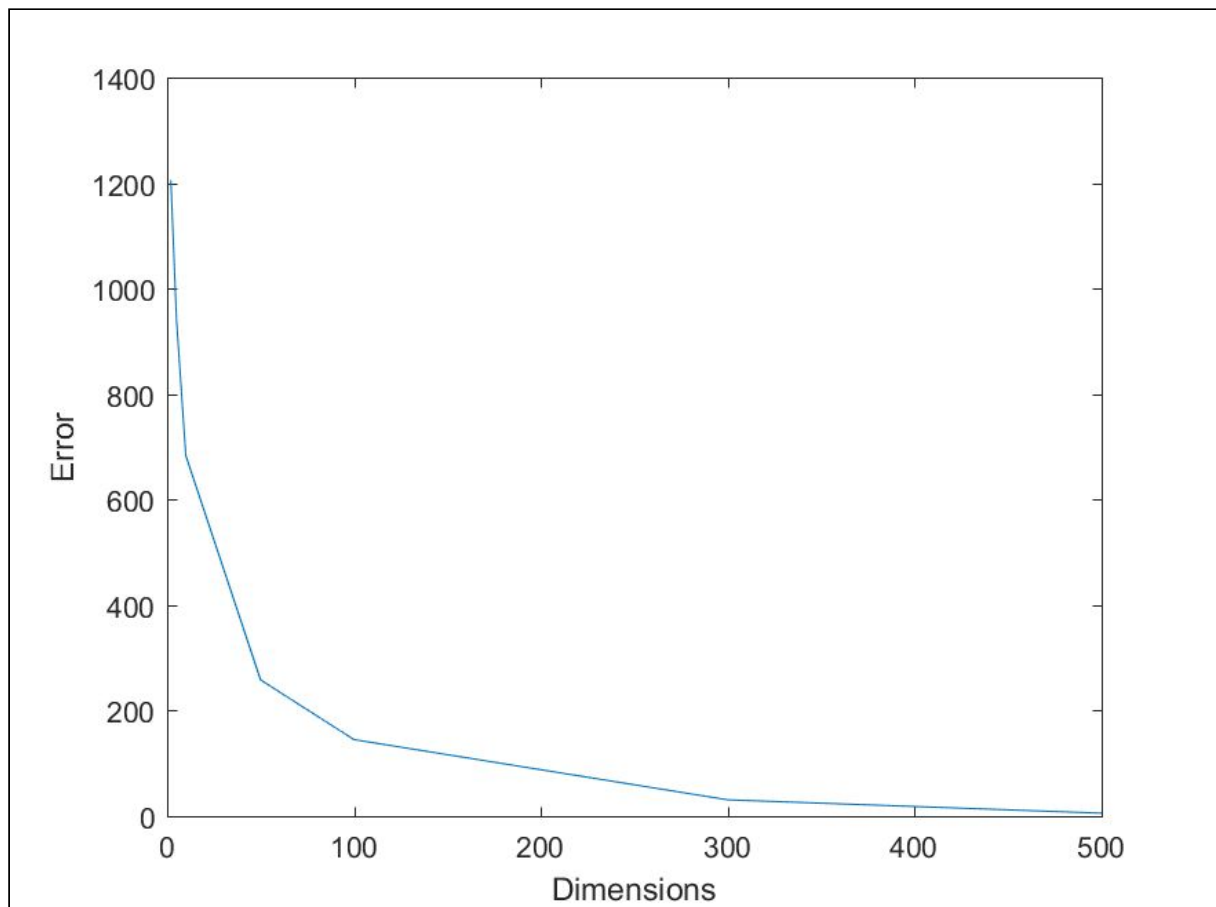
Si enlloc de plotejar les bases utilitzant la funció *imagesc* i utilitzem la funció *surf*, el que obtindrem serà la projecció d’aquestes bases en 3D, i en podrem veure millor els valors de la intensitat per a cada base.

3. Al reduir la dimensionalitat dels descriptors, que inicialment és 16384x1, ens desfem d'algunes de les dimensions d'aquests, provocant una pèrdua d'informació. Així doncs, com més dimensions reduïm, més informació perdem. Això ho podem veure representat en les següents imatges, on comencem reduint els descriptors a un nombre de dimensions molt baix, i l'anem incrementant. A mesura que incrementem les dimensions a les que reduïm els descriptors, veiem com la imatge es veu cada vegada més definida i menys borrosa, degut a que cada vegada es perd menys informació.



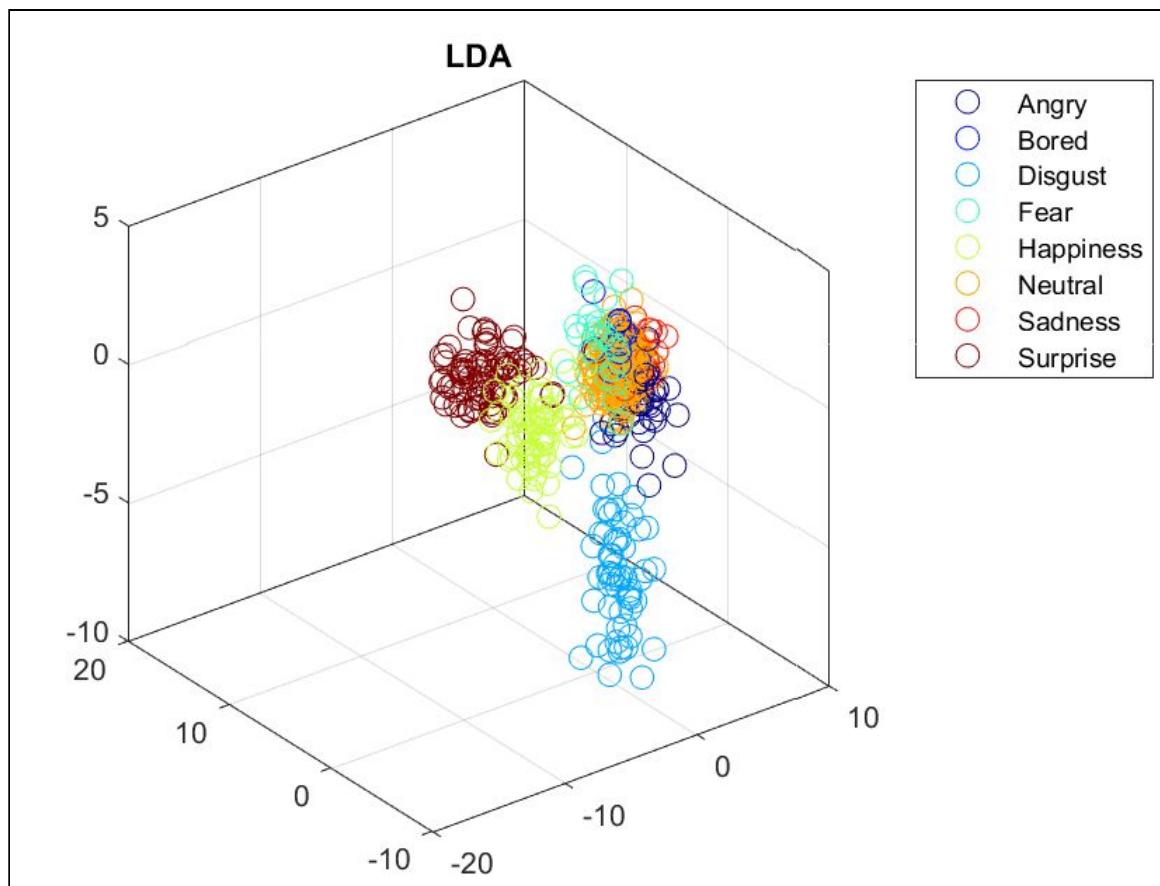
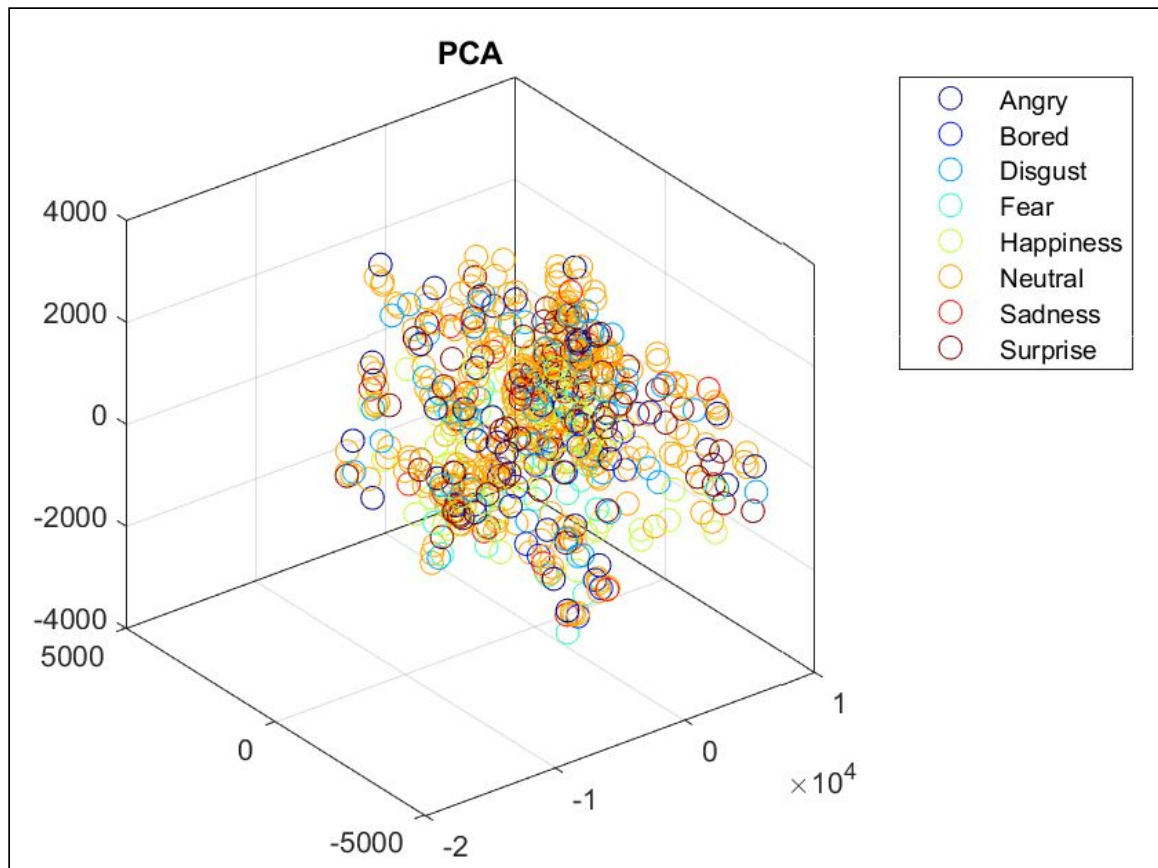


4. La gràfica que veiem a continuació ens mostra l'evolució de l'error en funció del nombre de dimensions al que reduïm els descriptors amb PCA. Com podem veure, representa una funció exponencial, i a partir de 100 dimensions l'error incrementa radicalment. Per aquest motiu, el nombre de dimensions al qual considerem que podem reduir sense perdre massa informació és 200.



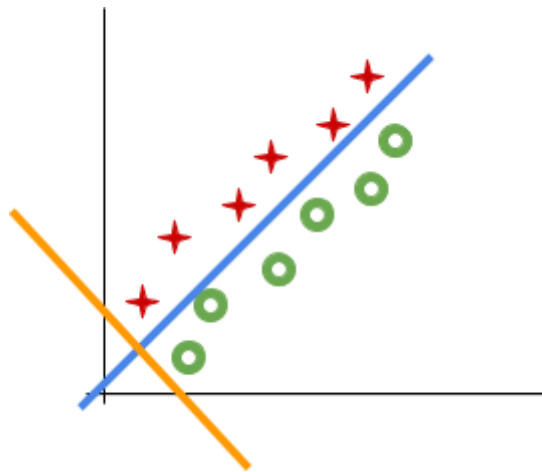
Podem dir que entre 0 i 50 dimensions la classificació de les imatges serà força errònia i poc fiable.

5. Per a obtenir els plots següents hem reduït la dimensió de les dades amb PCA i , en el cas de LDA, reduïm altra vegada les dimensions a 3 a partir de les dades projectades amb PCA.



La diferència de resultats amb LDA i PCA és deguda a que els dos mètodes funcionen de maneres totalment contràries:

- PCA busca la recta de màxima variança on projectar les dades amb mínima distància. Diguem que, tenint les dades que veiem a la gràfica que hi ha a continuació, PCA prescindiria de la dimensió representada en color taronja i es quedaria amb la que representa la màxima variança de les dades (color blau) i les projectaria en aquesta.
- En canvi, LDA busca projectar les dades en una dimensió reduïda on la separació entre les classes sigui màxima. Per tant, basant-nos en la mateixa gràfica, LDA prescindiria de la recta blava i es quedaria amb les dades projectades en la línia taronja, cosa que ja ens proporcionaria la classificació d'aquestes.



6. Agafem un mètode d'error comú per a visualitzar la diferència d'*accuracys* entre els mètodes. En la pràctica anterior vam concloure que un dels mètodes d'error que millor classificaven era el de correlació. Així, per a totes les proves següents hem utilitzat aquest mètode.

<i>accuracy</i>	2	5	10	50	100	300	500
PCA	0.2171	0.2844	0.4083	0.4709	0.4740	0.4847	0.4771
LDA	0.3777	0.5856	0.7125	0.7401	0.7018	0.7156	0.7110

Observem que el mètode de reducció de dimensionalitat LDA ens dona millors resultat que PCA amb una gran diferència. Això és degut a, tal i com s'ha explicat en el punt anterior, LDA fa una millor separació de les dades.

Per altra banda, es pot concloure que com menys dimensions tenim més informació perdem. Això es pot veure en que l'*accuracy* va augmentant a mesura que augmentem les dimensions. Tot i així, a partir de 10 dimensions l'*accuracy* es va mantenint força constant. D'aquesta forma, podríem passar de tenir unes dades en 500 dimensions a 50 o més de manera que, encara que perdem informació, els resultats de la classificació segueixen sent bons.

Per visualitzar-ho millor podem plotejar les diferents *accuracys* en funció de les dimensions, i veurem amb més claredat com es comporten constants a partir d'una certa reducció.

