2. Prove that for X, Y: $W^* = \underset{W \in O_d(\mathbb{R})}{\arg\min} \|WX - Y\|_F = UV^T$ with $U\Sigma V^T = SVD(YX^T)$

Minimizing $\|WX - Y\|_F \iff$ minimizing $\|WX - Y\|_F^2$

Let $X, Y \in \mathbb{R}^{d \times n}$: $\underset{W \in O_d(\mathbb{R})}{\min} \|WX - Y\|^2 = \underset{W \in O_d(\mathbb{R})}{\min} (WX^T)WX - 2(WX^T)Y + Y^TY$

$$= \underset{W \in O_d(\mathbb{R})}{\min} X^TX + Y^TY - 2YX^TW^T$$

because $W^TW = I$

$$= \underset{W \in O_d(\mathbb{R})}{\min} -2YX^TW^T$$

$$= \underset{W \in O_d(\mathbb{R})}{\max} YX^TW^T$$

$$= \underset{W \in O_d(\mathbb{R})}{\max} \Sigma V^TW^TU$$

$V^T, W^T, U \in O_d(\mathbb{R})^3 \Rightarrow V^TW^TU \in O_d(\mathbb{R})$

we should then choose $V^TW^TU = I_{d \times d} \Rightarrow W = UV^T$

4. Which loss did you use?

I used a categorical cross-entropy loss function because we have a multiclass classification problem

$$H(p, q) = -\sum_{i=1}^{5} y_i \log(y_i')$$ with y_i true probability distribution and y_i' predicted one

Be creative, use another encoder

I used XGBoost classifier, however it did not seem to add much value when submitting to Kaggle

3. Training and dev error

Training set accuracy: 0.48
Dev set accuracy: 0.42