# Introduction of Deepfake Representation with Multilinear Regression paper

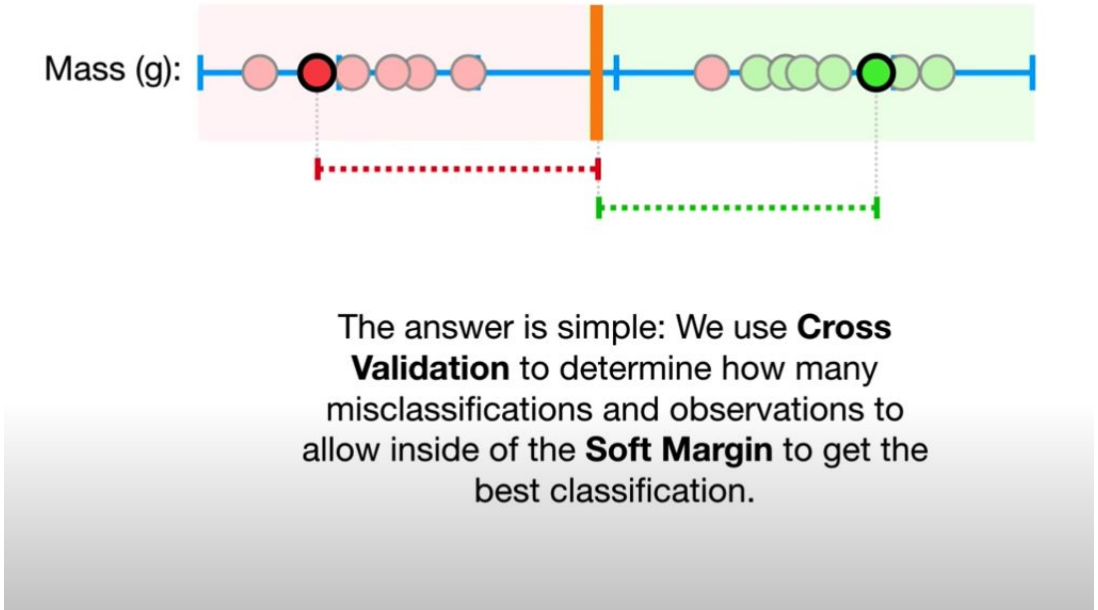
Yunting Chiu

# Dateset

- Dataset: FaceForensics videos
- experiment on **images** manipulated by DeepFake technique.
- Each 30 sec. extract 1 frame, total 7 frames for each video using OpenCV
- Using pretrained **dlib face detector** for detecting **outer facial landmarks**
- Training videos: 720 * 7(frames) = 5040 (images)
- Validation set : 140 videos
- Testing videos: 140 videos
- wd: Datasets/manipulated_sequences/Deepfakes/raw/videos

# Support Vector Machine (SVM)

- Binary Classification
- Pros: effective in high dimensional spaces
- Cons: long training time for large datasets




PC: StatQuest

**Algorithm 1** DeepFake Detection Algorithm

**Input** : $\mathbf{D}_{\text{real}}, \mathbf{D}_{\text{fake}}$ were centered by subtracting the mean of the real training data,

(1) Preprocessing and data tensor organization:
$[\mathbf{U}_{\text{real}}, \mathbf{S}_{\text{real}}, \mathbf{V}_{\text{real}}] \Leftarrow \text{svd}(\mathbf{D}_{\text{real}})$
$[\mathbf{U}_{\text{fake}}, \mathbf{S}_{\text{fake}}, \mathbf{V}_{\text{fake}}] \Leftarrow \text{svd}(\mathbf{D}_{\text{fake}})$
$\mathcal{D}(:,:,1) = [\mathbf{U}_{\text{real}}\mathbf{S}_{\text{real}}]$
$\mathcal{D}(:,:,2) = [\mathbf{U}_{\text{fake}}\mathbf{S}_{\text{fake}}]$

(2) Training data decomposition:
$\mathcal{T} \times_2 \mathbf{U}_{\text{f}} \times_3 \mathbf{U}_{\text{c}} \Leftarrow M\text{-mode SVD}(\mathcal{D})$

(3) Embed the class representations in the higher three dimensional space and set the third coordinate of the real and fake class to +1 and −1 respectively. Hence, $\mathbf{U}_{\text{c}} \in \mathbb{R}^{2\times2}$ now has dimensionality $\mathbb{R}^{2\times3}$. Normalize the rows of $\mathbf{U}_{\text{c}}$ to have length 1.

(4) Computer the extended core

$$\mathcal{T} := \mathcal{D} \times_2 \mathbf{U}_{\text{f}}^{\text{T}} \times_3 \mathbf{U}_{\text{c}}^{\dagger} \tag{16}$$

(5) Centering: validation and test data is centered by subtracting the mean of the real training data.

(6) Test data decomposition of a centered $\mathbf{d}_{\text{test}}$ :

$$\mathbf{d}_{\text{test}} \simeq \mathcal{T} \times_2 \mathbf{r}_{\text{f}}^{\text{T}} \times_3 \mathbf{r}_{\text{c}}^{\text{T}} \Leftarrow \text{Multilinear Projection}(\mathcal{T}, \mathbf{d}_{\text{test}})$$

(7) Finding linear SVM decision boundaries using validation set

(8) classifying all $\mathbf{d}_{\text{test}} \in$ test set

# Feature & Result

- Evaluate the decision boundaries
- Frames in the range 2980-5000, the accuracy is around 82 %. Otherwise, could be considered noise
- The range 2980-5000 class is more separable linearly.

| $\mathbf{U}_c \in \mathbb{R}^{2\times3}, \mathbf{U}_f \in \mathbb{R}^{5040\times R_f}$ | | | | | | | $R_f$ | **TN/140** | **TP/140** | **ACC** |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 721 | 1441 | 2161 | 2881 | 3601 | 4321 | 1-5040 | 98 | 101 | 0.7107 |
| | | | | | | | 1-720 | 107 | 93 | 0.7143 |
| | | | | | | | 721-2160 | 100 | 90 | 0.6786 |
| | | | | | | | 2161-3600 | 112 | 122 | 0.8000 |
| | | | | | | | 3601-5040 | 113 | 98 | 0.7536 |
| | | | | | | | 4321-5040 | 111 | 89 | 0.7143 |
| | | | | | | | 2161-5040 | 117 | 112 | 0.8179 |
| | | | | | | | **2980-5000** | **118** | **112** | **0.8214** |
| | | | | | | | 2881-5040 | 117 | 103 | 0.7857 |