

# Analyses of Convolution Neural Networks for automatic tagging of music tracks

Aravind Sankaran

RWTH Aachen

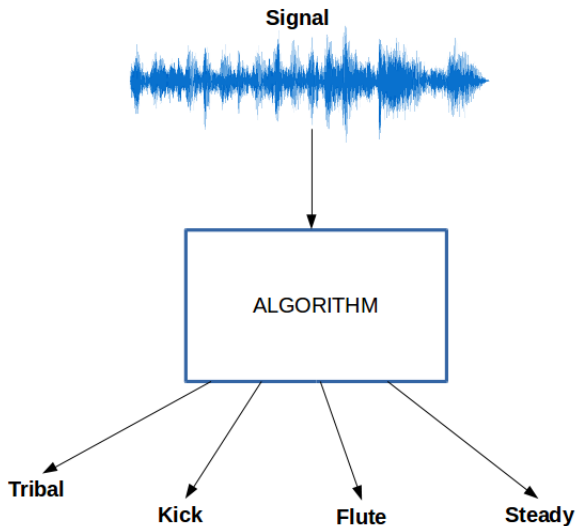
*aravind.sankaran@rwth-aachen.de*

April 6, 2017

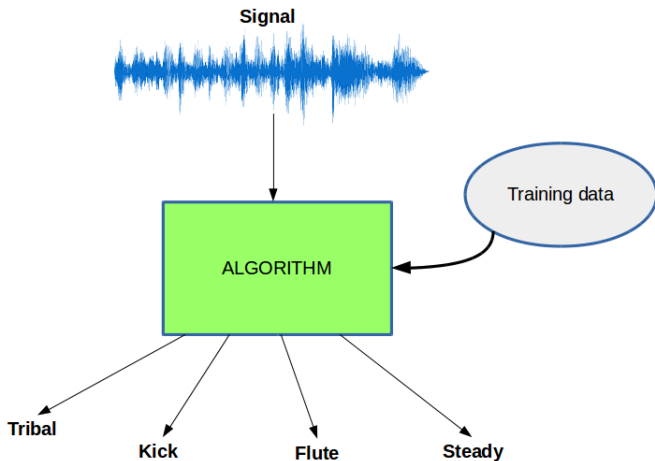
# Acknowledgement

- Prof. Paolo Bientinesi
- Prof. Marco Alunno

# AIM



# AIM



# APPLICATION

## User specific recommendation system?

→ Do you have to create a dataset with ground-truth?

# APPLICATION

## User specific recommendation system?

- **Do you have to create a dataset with ground-truth?**  
→ *Lets look at some solutions where you **don't** have to..*

### Collaborative filtering :

- Exploits social trends
- No information from audio content is used
- Cold-Start Problem

# APPLICATION

## User specific recommendation system?

- **Do you have to create a dataset with ground-truth?**
  - *Lets look at some solutions where you **don't** have to..*

### Collaborative + Content-based :

- Gather training data by crowd sourcing
- User specific recommendations by filtering popular tags

# APPLICATION

User specific recommendation system?

→ **When** do you have to create a dataset with ground-truth?



# APPLICATION

## Recommendation system for **experts**?

- **When** do you have to create a dataset with ground-truth?
- *Lets look from an artist's point of view ..*

# PIPELINE



# PIPELINE



## Inputs :

- Digital Signal (.mp3, .wav)
- Sheets of musical notes

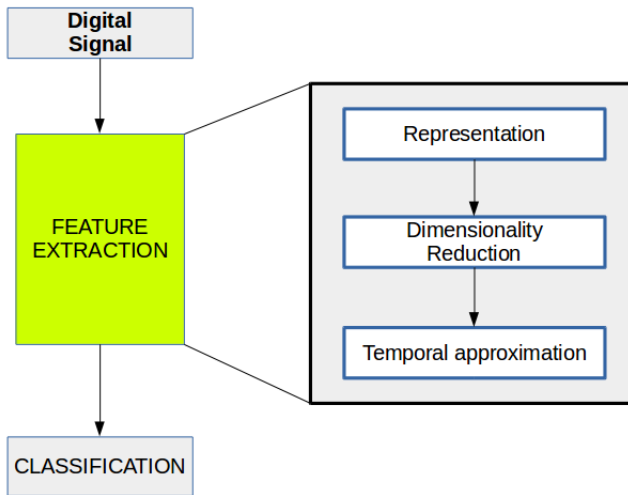
# PIPELINE



## Feature Extraction :

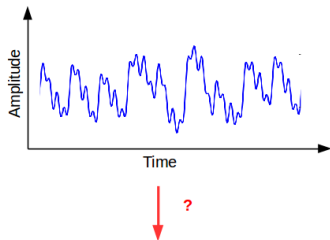
- $\mathbb{R}^N \rightarrow \mathbb{R}^T \quad T < N$
- **Organized :**  
Encode information about discriminants
- **Robust :**  
Transformation should be well posed

# PIPELINE

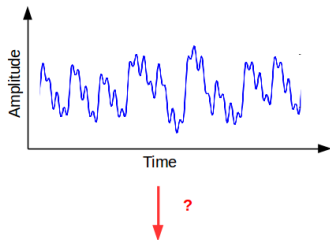


# REPRESENTATION

$$\mathbf{a} = (a_1, a_2, \dots, a_N) = a_1 \mathbf{e}_1 + \dots + a_N \mathbf{e}_N$$



# REPRESENTATION

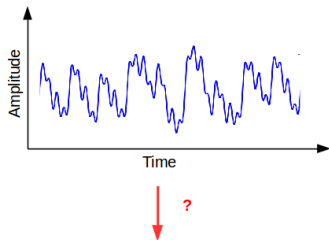


$$\mathbf{a} = (a_1, a_2, \dots, a_N) = a_1 \mathbf{e}_1 + \dots + a_N \mathbf{e}_N$$

## Basis

A group of vectors forms a basis of a vector space  $\mathbb{V}$  if every vector in  $\mathbb{V}$  can be represented as a linear combination of the basis vectors

# REPRESENTATION

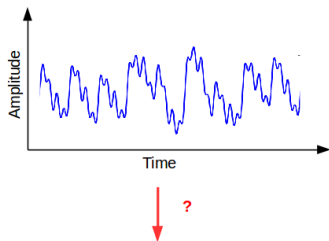


$$\mathbf{a} = a_1 \mathbf{e}_1 + \dots a_N \mathbf{e}_N = \mathbb{I}_N \mathbf{a}$$

$$\mathbf{a} = c_1 \mathbf{q}_1 + \dots c_M \mathbf{q}_M = \mathbf{Q} \mathbf{c}$$



# REPRESENTATION



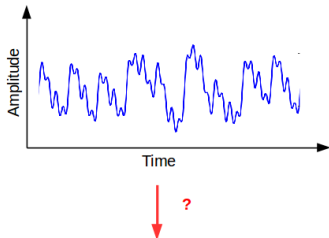
$$\mathbf{a} = a_1 \mathbf{e}_1 + \dots a_N \mathbf{e}_N = \mathbb{I}_N \mathbf{a}$$

$$\mathbf{a} = c_1 \mathbf{q}_1 + \dots c_M \mathbf{q}_M = \mathbf{Q} \mathbf{c}$$

$$\mathbf{Q}^{-1} \mathbf{a} = \mathbf{c} \quad \mathbf{Q}^{-1} \in \mathbb{R}^{M \times N}$$

# REPRESENTATION

$$\mathbf{Q}^{-1}\mathbf{a} = \mathbf{c} \quad \mathbf{Q}^{-1} \in \mathbb{R}^{M \times N}$$

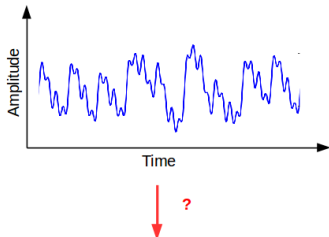


## Exponential Fourier Theorem

Complex exponentials which are functions of frequencies form basis for *periodic* function

# REPRESENTATION

$$\mathbf{Q}^{-1}\mathbf{a} = \mathbf{c} \quad \mathbf{Q}^{-1} \in \mathbb{R}^{M \times N}$$



## Exponential Fourier Theorem

Complex exponentials which are functions of frequencies form basis for *periodic* function

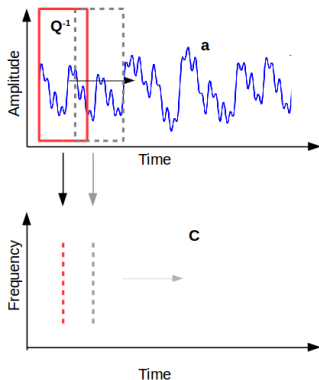
## Fourier Transform

Application of *Fourier Theorem* for general signals.

$$\mathbf{Q}^{-1}[i] = \mathbf{e}^{-j\omega t} \quad i \in \{0, 1.., M\}$$

# REPRESENTATION

$$\mathbf{Q}^{-1}\mathbf{a} = \mathbf{c} \quad \mathbf{Q}^{-1} \in \mathbb{R}^{M \times N}$$

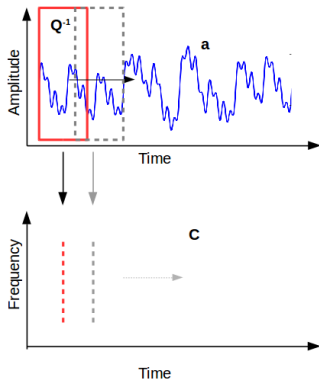


## Fourier Transform

Application of *Fourier Theorem* for general signals.

$$\mathbf{Q}^{-1}[i] = \mathbf{e}^{-j\omega t} \quad i \in \{0, 1.., M\}$$

# REPRESENTATION



$$\mathbf{Q}^{-1}\mathbf{a} = \mathbf{c} \quad \mathbf{Q}^{-1} \in \mathbb{R}^{M \times N}$$

## Fourier Transform

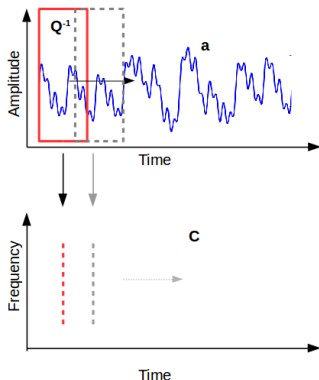
Application of *Fourier Theorem* for general signals.

$$\mathbf{Q}^{-1}[i] = \mathbf{e}^{-j\omega t} \quad i \in \{0, 1, \dots, M\}$$

## Short-time Fourier Transform

$$\mathbf{C} = \mathbf{a} \star \mathbf{Q}^{-1} \quad \mathbf{C} \in \mathbb{C}^{M \times P}$$

# REPRESENTATION



## Short-time Fourier Transform

$$\mathbf{C} = \mathbf{a} \star \mathbf{Q}^{-1} \quad \mathbf{C} \in \mathbb{C}^{M \times P}$$

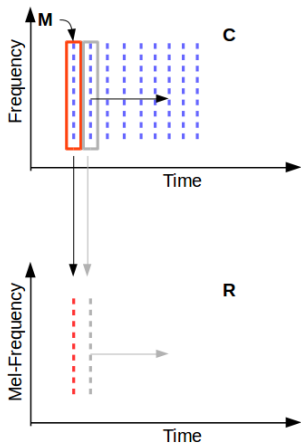
## Fast Fourier Transform

Faster version of STFT that exploits the symmetry of sinusoids.

$$\text{STFT} : O(N^2)$$

$$\text{FFT} : O(N \log N)$$

# REPRESENTATION

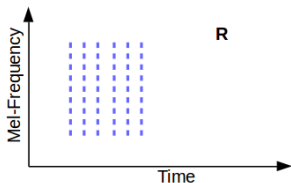


$$C = a \star Q^{-1} \quad C \in \mathbb{C}^{M \times P}$$

## Spectrogram representations

- **Mel Spec :**  
 $R = M.C \quad \forall M \in \mathbb{R}^{R \times M}$
- **Chromagram :**  
 $R = C.M_C$
- **Tempogram :**  
 $R = C \star M_T$

# DIMENSIONALITY REDUCTION



$$\mathbf{C} = \mathbf{a} \star \mathbf{Q}^{-1} \quad \mathbf{C} \in \mathbb{C}^{M \times P}$$

$$\mathbf{R} = \mathbf{C} \star \mathbf{M} \quad \mathbf{R} \in \mathbb{R}^{R \times P}$$

## Principal Component Analysis

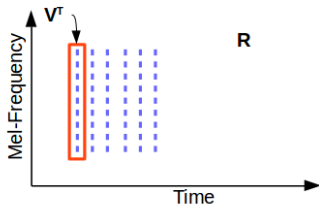
Represent  $\mathbf{R}$  in a basis that is a function of variance in the information.



# DIMENSIONALITY REDUCTION

$$\mathbf{C} = \mathbf{a} \star \mathbf{Q}^{-1} \quad \mathbf{C} \in \mathbb{C}^{M \times P}$$

$$\mathbf{R} = \mathbf{C} \star \mathbf{M} \quad \mathbf{R} \in \mathbb{R}^{R \times P}$$



## Principal Component Analysis

Represent  $\mathbf{R}$  in a basis that is a function of variance in the information.

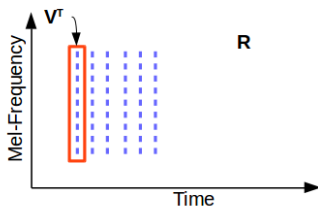
$$\hat{\mathbf{R}} = \text{Center}(\mathbf{R})$$

$$\Sigma = \frac{1}{P} \hat{\mathbf{R}} \hat{\mathbf{R}}^T$$

$$\mathbf{V} \Lambda \mathbf{V}^T = \Sigma$$

$$\mathbf{X} = \text{Truncate}(\mathbf{V}^T) \hat{\mathbf{R}}$$

# DIMENSIONALITY REDUCTION



$$\mathbf{C} = \mathbf{a} \star \mathbf{Q}^{-1} \quad \mathbf{C} \in \mathbb{C}^{M \times P}$$

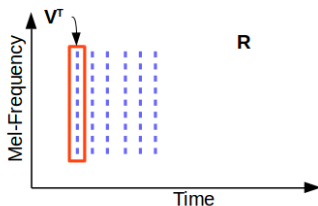
$$\mathbf{R} = \mathbf{C} \star \mathbf{M} \quad \mathbf{R} \in \mathbb{R}^{R \times P}$$

## Mel-Frequency Cepstral Coefficients

Basis functions of *principal components* of log spectra are very similar to *cosine transform*

$$\mathbf{V}^T[i] = \cos(\omega t) \quad i \in \{0, 1, \dots, T\}$$

# DIMENSIONALITY REDUCTION



$$\mathbf{C} = \mathbf{a} \star \mathbf{Q}^{-1} \quad \mathbf{C} \in \mathbb{C}^{M \times P}$$

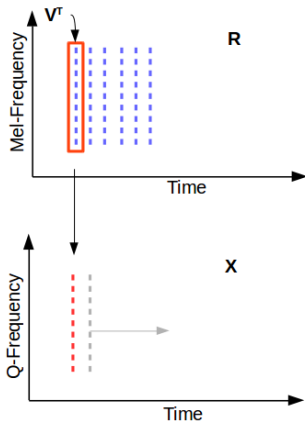
$$\mathbf{R} = \mathbf{C} \star \mathbf{M} \quad \mathbf{R} \in \mathbb{R}^{R \times P}$$

## Mel-Frequency Cepstral Coefficients

Basis functions of *principal components* of log spectra are very similar to *cosine transform*

$$\mathbf{V}^T[i] = \cos(\omega t) \quad i \in \{0, 1, \dots, T\}$$

# DIMENSIONALITY REDUCTION



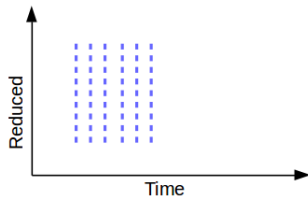
$$\begin{aligned} C &= a \star Q^{-1} & C &\in \mathbb{C}^{M \times P} \\ R &= C \star M & R &\in \mathbb{R}^{R \times P} \\ X &= R \star V^T & X &\in \mathbb{R}^{T \times P} \end{aligned}$$

## Mel-Frequency Cepstral Coefficients

Basis functions of *principal components* of log spectra are very similar to *cosine transform*

$$V^T[i] = \cos(\omega t) \quad i \in \{0, 1, \dots, T\}$$

# TEMPORAL APPROXIMATION



$$\mathbf{C} = \mathbf{a} \star \mathbf{Q}^{-1}$$

$$\mathbf{R} = \mathbf{C} \star \mathbf{M}$$

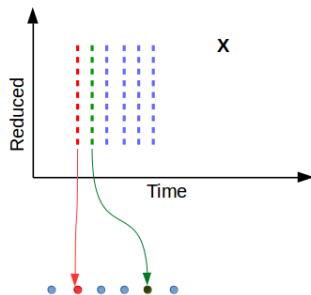
$$\mathbf{X} = \mathbf{R} \star \mathbf{V}^T$$

$$\mathbf{C} \in \mathbb{C}^{M \times P}$$

$$\mathbf{R} \in \mathbb{R}^{R \times P}$$

$$\mathbf{X} \in \mathbb{R}^{T \times P}$$

# TEMPORAL APPROXIMATION



$$\begin{aligned} \mathbf{C} &= \mathbf{a} \star \mathbf{Q}^{-1} & \mathbf{C} &\in \mathbb{C}^{M \times P} \\ \mathbf{R} &= \mathbf{C} \star \mathbf{M} & \mathbf{R} &\in \mathbb{R}^{R \times P} \\ \mathbf{X} &= \mathbf{R} \star \mathbf{V}^T & \mathbf{X} &\in \mathbb{R}^{T \times P} \end{aligned}$$

## Bag Of Frames

- Assign each column of  $\mathbf{X}$  to the nearest of  $K$  clusters.
- Count the number of assignments to each of the  $K$  clusters.
- The resulting feature is of dimension  $K$

# TEMPORAL APPROXIMATION

## K - Means[1] :

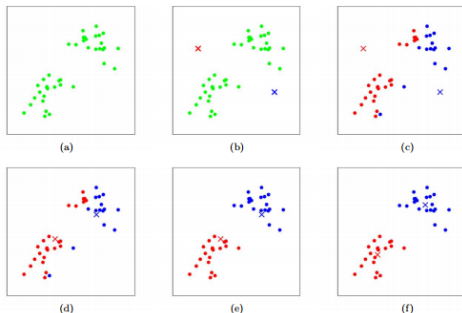


Figure 1: K-means algorithm. Training examples are shown as dots, and cluster centroids are shown as crosses. (a) Original dataset. (b) Random initial cluster centroids. (c-f) Illustration of running two iterations of k-means. In each iteration, we assign each training example to the closest cluster centroid (shown by "painting" the training examples the same color as the cluster centroid to which is assigned); then we move each cluster centroid to the mean of the points assigned to it. Images courtesy of Michael Jordan.

# CLASSIFICATION

**Feature Extraction :**  $\mathbb{R}^{N_f} \rightarrow \mathbb{R}^K$

$$\mathbf{C} = \mathbf{a} \star \mathbf{Q}^{-1} \quad \mathbf{C} \in \mathbb{C}^{M \times P}$$

$$\mathbf{R} = \mathbf{C} \star \mathbf{M} \quad \mathbf{R} \in \mathbb{R}^{R \times P}$$

$$\mathbf{X} = \mathbf{R} \star \mathbf{V}^T \quad \mathbf{X} \in \mathbb{R}^{T \times P}$$

$$\mathbf{f} = \textit{Temporal\_Approx}(\mathbf{X}) \quad \mathbf{f} \in \mathbb{R}^K$$

**Classification :**  $\mathbb{R}^K \rightarrow \mathbb{R}^L$

$$\mathbf{y} = \Phi(\mathbf{f}) \quad \mathbf{y} \in \mathbb{R}^L$$



# CLASSIFICATION

Single-layer perceptron :

$$\mathbf{y} = \mathbf{W}\mathbf{f} \quad \mathbf{W} \in \mathbb{R}^{L \times K}$$

$$\mathbf{FE} : \mathbb{R}^{N_f} \rightarrow \mathbb{R}^K$$

$$\mathbf{C} = \mathbf{a} \star \mathbf{Q}^{-1} \quad \mathbf{C} \in \mathbb{C}^{M \times P}$$

$$\mathbf{R} = \mathbf{C} \star \mathbf{M} \quad \mathbf{R} \in \mathbb{R}^{R \times P}$$

$$\mathbf{X} = \mathbf{R} \star \mathbf{V}^T \quad \mathbf{X} \in \mathbb{R}^{T \times P}$$

$$\mathbf{f} = T(\mathbf{X}) \quad \mathbf{f} \in \mathbb{R}^K$$

$$\text{Classification} : \mathbb{R}^K \rightarrow \mathbb{R}^L$$

$$\mathbf{y} = \Phi(\mathbf{f}) \quad \mathbf{y} \in \mathbb{R}^L$$

# CLASSIFICATION

Single-layer perceptron :

$$\mathbf{y} = \mathbf{W}\mathbf{f} \quad \mathbf{W} \in \mathbb{R}^{L \times K}$$

Solve for  $\mathbf{W}$  with the training data

$$\mathbf{FE} : \mathbb{R}^{N_f} \rightarrow \mathbb{R}^K$$

$$\mathbf{C} = \mathbf{a} \star \mathbf{Q}^{-1} \quad \mathbf{C} \in \mathbb{C}^{M \times P}$$

$$\mathbf{R} = \mathbf{C} \star \mathbf{M} \quad \mathbf{R} \in \mathbb{R}^{R \times P}$$

$$\mathbf{X} = \mathbf{R} \star \mathbf{V}^T \quad \mathbf{X} \in \mathbb{R}^{T \times P}$$

$$\mathbf{f} = T(\mathbf{X}) \quad \mathbf{f} \in \mathbb{R}^K$$

$$\text{Classification} : \mathbb{R}^K \rightarrow \mathbb{R}^L$$

$$\mathbf{y} = \Phi(\mathbf{f}) \quad \mathbf{y} \in \mathbb{R}^L$$

# CLASSIFICATION

Two-layer perceptron :

$$\mathbf{y} = \sigma(\mathbf{W}_2 \text{ReLU}(\mathbf{W}_1 \mathbf{f}))$$

$$\mathbf{W}_2 \in \mathbb{R}^{L \times H} \quad \mathbf{W}_1 \in \mathbb{R}^{H \times K}$$

$$\mathbf{FE} : \mathbb{R}^{N_f} \rightarrow \mathbb{R}^K$$

$$\mathbf{C} = \mathbf{a} \star \mathbf{Q}^{-1} \quad \mathbf{C} \in \mathbb{C}^{M \times P}$$

$$\mathbf{R} = \mathbf{C} \star \mathbf{M} \quad \mathbf{R} \in \mathbb{R}^{R \times P}$$

$$\mathbf{X} = \mathbf{R} \star \mathbf{V}^T \quad \mathbf{X} \in \mathbb{R}^{T \times P}$$

$$\mathbf{f} = T(\mathbf{X}) \quad \mathbf{f} \in \mathbb{R}^K$$

$$\text{Classification} : \mathbb{R}^K \rightarrow \mathbb{R}^L$$

$$\mathbf{y} = \Phi(\mathbf{f}) \quad \mathbf{y} \in \mathbb{R}^L$$

# CLASSIFICATION

$$\mathbf{y} = \sigma(\mathbf{W}_2 \text{ReLU}(\mathbf{W}_1 \mathbf{f}))$$

**Training :**

- $E = \text{loss}(\mathbf{y}, \mathbf{t})$

$$\mathbf{FE} : \mathbb{R}^{N_f} \rightarrow \mathbb{R}^K$$

$$\mathbf{C} = \mathbf{a} \star \mathbf{Q}^{-1} \quad \mathbf{C} \in \mathbb{C}^{M \times P}$$

$$\mathbf{R} = \mathbf{C} \star \mathbf{M} \quad \mathbf{R} \in \mathbb{R}^{R \times P}$$

$$\mathbf{X} = \mathbf{R} \star \mathbf{V}^T \quad \mathbf{X} \in \mathbb{R}^{T \times P}$$

$$\mathbf{f} = T(\mathbf{X}) \quad \mathbf{f} \in \mathbb{R}^K$$

$$\text{Classification} : \mathbb{R}^K \rightarrow \mathbb{R}^L$$

$$\mathbf{y} = \sigma(\mathbf{W}_2 \text{ReLU}(\mathbf{W}_1 \mathbf{f})) \quad \mathbf{y} \in \mathbb{R}^L$$

# CLASSIFICATION

$$\mathbf{y} = \sigma(\mathbf{W}_2 \text{ReLU}(\mathbf{W}_1 \mathbf{f}))$$

Training :

- $E = \text{loss}(\mathbf{y}, \mathbf{t})$
- $\frac{\partial E}{\partial \mathbf{W}_2} = \frac{\partial E}{\partial \mathbf{y}} \frac{\partial \mathbf{y}}{\partial \sigma} \frac{\partial \sigma}{\partial \mathbf{W}_2}$
- $\frac{\partial E}{\partial \mathbf{W}_1} = \frac{\partial E}{\partial \mathbf{y}} \frac{\partial \mathbf{y}}{\partial \sigma} \frac{\partial \sigma}{\partial \text{ReLU}} \frac{\partial \text{ReLU}}{\partial \mathbf{W}_1}$

$$\mathbf{FE} : \mathbb{R}^{N_f} \rightarrow \mathbb{R}^K$$

$$\mathbf{C} = \mathbf{a} \star \mathbf{Q}^{-1} \quad \mathbf{C} \in \mathbb{C}^{M \times P}$$

$$\mathbf{R} = \mathbf{C} \star \mathbf{M} \quad \mathbf{R} \in \mathbb{R}^{R \times P}$$

$$\mathbf{X} = \mathbf{R} \star \mathbf{V}^T \quad \mathbf{X} \in \mathbb{R}^{T \times P}$$

$$\mathbf{f} = T(\mathbf{X}) \quad \mathbf{f} \in \mathbb{R}^K$$

$$\text{Classification} : \mathbb{R}^K \rightarrow \mathbb{R}^L$$

$$\mathbf{y} = \sigma(\mathbf{W}_2 \text{ReLU}(\mathbf{W}_1 \mathbf{f})) \quad \mathbf{y} \in \mathbb{R}^L$$

# CLASSIFICATION

$$\mathbf{y} = \sigma(\mathbf{W}_2 \text{ReLU}(\mathbf{W}_1 \mathbf{f}))$$

**Training :**

- $E = \text{loss}(\mathbf{y}, \mathbf{t})$
- $\frac{\partial E}{\partial \mathbf{W}_2} = \frac{\partial E}{\partial \mathbf{y}} \frac{\partial \mathbf{y}}{\partial \sigma} \frac{\partial \sigma}{\partial \mathbf{W}_2}$
- $\frac{\partial E}{\partial \mathbf{W}_1} = \frac{\partial E}{\partial \mathbf{y}} \frac{\partial \mathbf{y}}{\partial \sigma} \frac{\partial \sigma}{\partial \text{ReLU}} \frac{\partial \text{ReLU}}{\partial \mathbf{W}_1}$
- $\mathbf{W}_1 \leftarrow \text{update}(\mathbf{W}_1, \frac{\partial E}{\partial \mathbf{W}_1})$
- $\mathbf{W}_2 \leftarrow \text{update}(\mathbf{W}_2, \frac{\partial E}{\partial \mathbf{W}_2})$

$$\mathbf{FE} : \mathbb{R}^{N_f} \rightarrow \mathbb{R}^K$$

$$\mathbf{C} = \mathbf{a} \star \mathbf{Q}^{-1} \quad \mathbf{C} \in \mathbb{C}^{M \times P}$$

$$\mathbf{R} = \mathbf{C} \star \mathbf{M} \quad \mathbf{R} \in \mathbb{R}^{R \times P}$$

$$\mathbf{X} = \mathbf{R} \star \mathbf{V}^T \quad \mathbf{X} \in \mathbb{R}^{T \times P}$$

$$\mathbf{f} = T(\mathbf{X}) \quad \mathbf{f} \in \mathbb{R}^K$$

$$\text{Classification} : \mathbb{R}^K \rightarrow \mathbb{R}^L$$

$$\mathbf{y} = \sigma(\mathbf{W}_2 \text{ReLU}(\mathbf{W}_1 \mathbf{f})) \quad \mathbf{y} \in \mathbb{R}^L$$

## OVERVIEW

$$\mathbf{FE} : \mathbb{R}^{N_f} \rightarrow \mathbb{R}^K$$

$$\mathbf{C} = \mathbf{a} \star \mathbf{Q}^{-1} \quad \mathbf{C} \in \mathbb{C}^{M \times P}$$

$$\mathbf{R} = \mathbf{C} \star \mathbf{M} \quad \mathbf{R} \in \mathbb{R}^{R \times P}$$

$$\mathbf{X} = \mathbf{R} \star \mathbf{V}^T \quad \mathbf{X} \in \mathbb{R}^{T \times P}$$

$$\mathbf{f} = T(\mathbf{X}) \quad \mathbf{f} \in \mathbb{R}^K$$

$$\text{Classification} : \mathbb{R}^K \rightarrow \mathbb{R}^L$$

$$\mathbf{y} = \sigma(\mathbf{W}_2 \text{ReLU}(\mathbf{W}_1 \mathbf{f})) \quad \mathbf{y} \in \mathbb{R}^L$$

## SUPERVISING FEATURES

Recurrent Neural Network:

$$\mathbf{FE} : \mathbb{R}^{N_f} \rightarrow \mathbb{R}^K$$

$$\mathbf{C} = \mathbf{a} \star \mathbf{Q}^{-1} \quad \mathbf{C} \in \mathbb{C}^{M \times P}$$

$$\mathbf{R} = \mathbf{C} \star \mathbf{M} \quad \mathbf{R} \in \mathbb{R}^{R \times P}$$

$$\mathbf{X} = \mathbf{R} \star \mathbf{V}^T \quad \mathbf{X} \in \mathbb{R}^{T \times P}$$

$$\mathbf{f} = T(\mathbf{X}) \quad \mathbf{f} \in \mathbb{R}^K$$

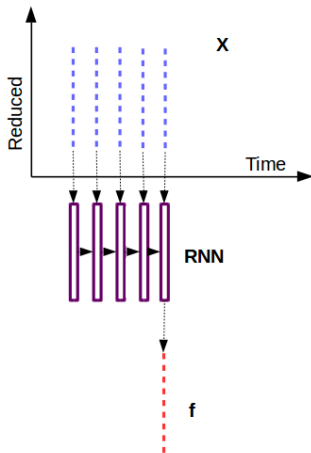
$$\text{Classification} : \mathbb{R}^K \rightarrow \mathbb{R}^L$$

$$\mathbf{y} = \sigma(\mathbf{W}_2 \text{ReLU}(\mathbf{W}_1 \mathbf{f})) \quad \mathbf{y} \in \mathbb{R}^L$$



# SUPERVISING FEATURES

## Recurrent Neural Network: (Sequence to One)



$$\mathbf{FE} : \mathbb{R}^{N_f} \rightarrow \mathbb{R}^K$$

$$\mathbf{C} = \mathbf{a} \star \mathbf{Q}^{-1} \quad \mathbf{C} \in \mathbb{C}^{M \times P}$$

$$\mathbf{R} = \mathbf{C} \star \mathbf{M} \quad \mathbf{R} \in \mathbb{R}^{R \times P}$$

$$\mathbf{X} = \mathbf{R} \star \mathbf{V}^T \quad \mathbf{X} \in \mathbb{R}^{T \times P}$$

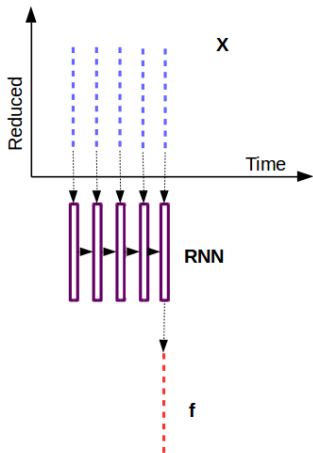
$$\mathbf{f} = \mathcal{T}(\mathbf{X}) \quad \mathbf{f} \in \mathbb{R}^K$$

$$\text{Classification} : \mathbb{R}^K \rightarrow \mathbb{R}^L$$

$$\mathbf{y} = \sigma(\mathbf{W}_2 \text{ReLU}(\mathbf{W}_1 \mathbf{f})) \quad \mathbf{y} \in \mathbb{R}^L$$

# SUPERVISING FEATURES

## Recurrent Neural Network: (Sequence to One)



### RNN

$$\mathbf{f} = \mathbf{W}_3 \mathbf{h}_P$$

$$\mathbf{h}_P = \Phi(\mathbf{X}[P], \mathbf{X}[P-1], \dots, \mathbf{X}[0])$$

# SUPERVISING FEATURES

## Recurrent Neural Network: (Sequence to One)

### RNN

$$\mathbf{f} = \mathbf{W}_3 \mathbf{h}_P \quad \mathbf{h}_P = \Phi(\mathbf{X}[P], \mathbf{X}[P-1], \dots, \mathbf{X}[0])$$

### Long Short Term Memory

$$\mathbf{f} = \mathbf{W}_3 \mathbf{h}_P$$

$$\mathbf{h}_p = \mathbf{o}_p \odot \sigma_h(\mathbf{c}_p)$$

$$\mathbf{c}_p = \mathbf{g}_p \odot \mathbf{c}_{p-1} + \mathbf{i}_p \odot \sigma_c(\mathbf{W}_c \mathbf{x}_p + \mathbf{U}_c \mathbf{h}_{p-1})$$

$$\mathbf{o}_p = \sigma(\mathbf{W}_o \mathbf{x}_p + \mathbf{U}_o \mathbf{h}_{p-1})$$

$$\mathbf{i}_p = \sigma(\mathbf{W}_i \mathbf{x}_p + \mathbf{U}_i \mathbf{h}_{p-1})$$

$$\mathbf{g}_p = \sigma(\mathbf{W}_g \mathbf{x}_p + \mathbf{U}_g \mathbf{h}_{p-1})$$

# SUPERVISING FEATURE

$$\mathbf{FE} : \mathbb{R}^{N_f} \rightarrow \mathbb{R}^K$$

$$\mathbf{C} = \mathbf{a} \star \mathbf{Q}^{-1} \quad \mathbf{C} \in \mathbb{C}^{M \times P}$$

$$\mathbf{R} = \mathbf{C} \star \mathbf{M} \quad \mathbf{R} \in \mathbb{R}^{R \times P}$$

$$\mathbf{X} = \mathbf{R} \star \mathbf{V}^T \quad \mathbf{X} \in \mathbb{R}^{T \times P}$$

$$\mathbf{f} = \text{LSTM}(\mathbf{X}) \quad \mathbf{f} \in \mathbb{R}^K$$

$$\text{Classification} : \mathbb{R}^K \rightarrow \mathbb{R}^L$$

$$\mathbf{y} = \sigma(\mathbf{W}_2 \text{ReLU}(\mathbf{W}_1 \mathbf{f})) \quad \mathbf{y} \in \mathbb{R}^L$$

# SUPERVISING FEATURE - Deep learning

## Convolution Neural Network :

$$\mathbf{FE} : \mathbb{R}^{N_f} \rightarrow \mathbb{R}^K$$

$$\mathbf{C} = \mathbf{a} \star \mathbf{Q}^{-1} \quad \mathbf{C} \in \mathbb{C}^{M \times P}$$

$$\mathbf{R} = \mathbf{C} \star \mathbf{M} \quad \mathbf{R} \in \mathbb{R}^{R \times P}$$

$$\mathbf{X} = \mathbf{R} \star \mathbf{W}_4 \quad \mathbf{X} \in \mathbb{R}^{T \times P}$$

$$\mathbf{f} = \text{LSTM}(\mathbf{X}) \quad \mathbf{f} \in \mathbb{R}^K$$

$$\text{Classification} : \mathbb{R}^K \rightarrow \mathbb{R}^L$$

$$\mathbf{y} = \sigma(\mathbf{W}_2 \text{ReLU}(\mathbf{W}_1 \mathbf{f})) \quad \mathbf{y} \in \mathbb{R}^L$$

# Overview

## 1 First Section

### ■ Subsection Example

## 2 Second Section

# Paragraphs of Text

Sed iaculis dapibus gravida. Morbi sed tortor erat, nec interdum arcu. Sed id lorem lectus. Quisque viverra augue id sem ornare non aliquam nibh tristique. Aenean in ligula nisl. Nulla sed tellus ipsum. Donec vestibulum ligula non lorem vulputate fermentum accumsan neque mollis.

Sed diam enim, sagittis nec condimentum sit amet, ullamcorper sit amet libero. Aliquam vel dui orci, a porta odio. Nullam id suscipit ipsum. Aenean lobortis commodo sem, ut commodo leo gravida vitae. Pellentesque vehicula ante iaculis arcu pretium rutrum eget sit amet purus. Integer ornare nulla quis neque ultrices lobortis. Vestibulum ultrices tincidunt libero, quis commodo erat ullamcorper id.

# Bullet Points

- Lorem ipsum dolor sit amet, consectetur adipiscing elit
- Aliquam blandit faucibus nisi, sit amet dapibus enim tempus eu
- Nulla commodo, erat quis gravida posuere, elit lacus lobortis est, quis porttitor odio mauris at libero
- Nam cursus est eget velit posuere pellentesque
- Vestibulum faucibus velit a augue condimentum quis convallis nulla gravida



# Blocks of Highlighted Text

## Block 1

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Integer lectus nisl, ultricies in feugiat rutrum, porttitor sit amet augue. Aliquam ut tortor mauris. Sed volutpat ante purus, quis accumsan dolor.

## Block 2

Pellentesque sed tellus purus. Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos himenaeos. Vestibulum quis magna at risus dictum tempor eu vitae velit.

## Block 3

Suspendisse tincidunt sagittis gravida. Curabitur condimentum, enim sed venenatis rutrum, ipsum neque consectetur orci, sed blandit justo nisi ac lacus.

# Multiple Columns

## Heading

- 1 Statement
- 2 Explanation
- 3 Example

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Integer lectus nisl, ultricies in feugiat rutrum, porttitor sit amet augue. Aliquam ut tortor mauris. Sed volutpat ante purus, quis accumsan dolor.

# Table

Treatments	Response 1	Response 2
Treatment 1	0.0003262	0.562
Treatment 2	0.0015681	0.910
Treatment 3	0.0009271	0.296

Table : Table caption

# Theorem

Theorem (Mass–energy equivalence)

$$E = mc^2$$

# Verbatim

## Example (Theorem Slide Code)

```
\begin{frame}  
\frametitle{Theorem}  
\begin{theorem}[Mass--energy equivalence]  
$E = mc^2$  
\end{theorem}  
\end{frame}
```

# Figure

Uncomment the code on this slide to include your own image from the same directory as the template .TeX file.

# Citation

An example of the `\cite` command to cite within the presentation:

This statement requires citation [Smith, 2012].

# References



John Smith (2012)

Title of the publication

*Journal Name* 12(3), 45 – 678.



# The End