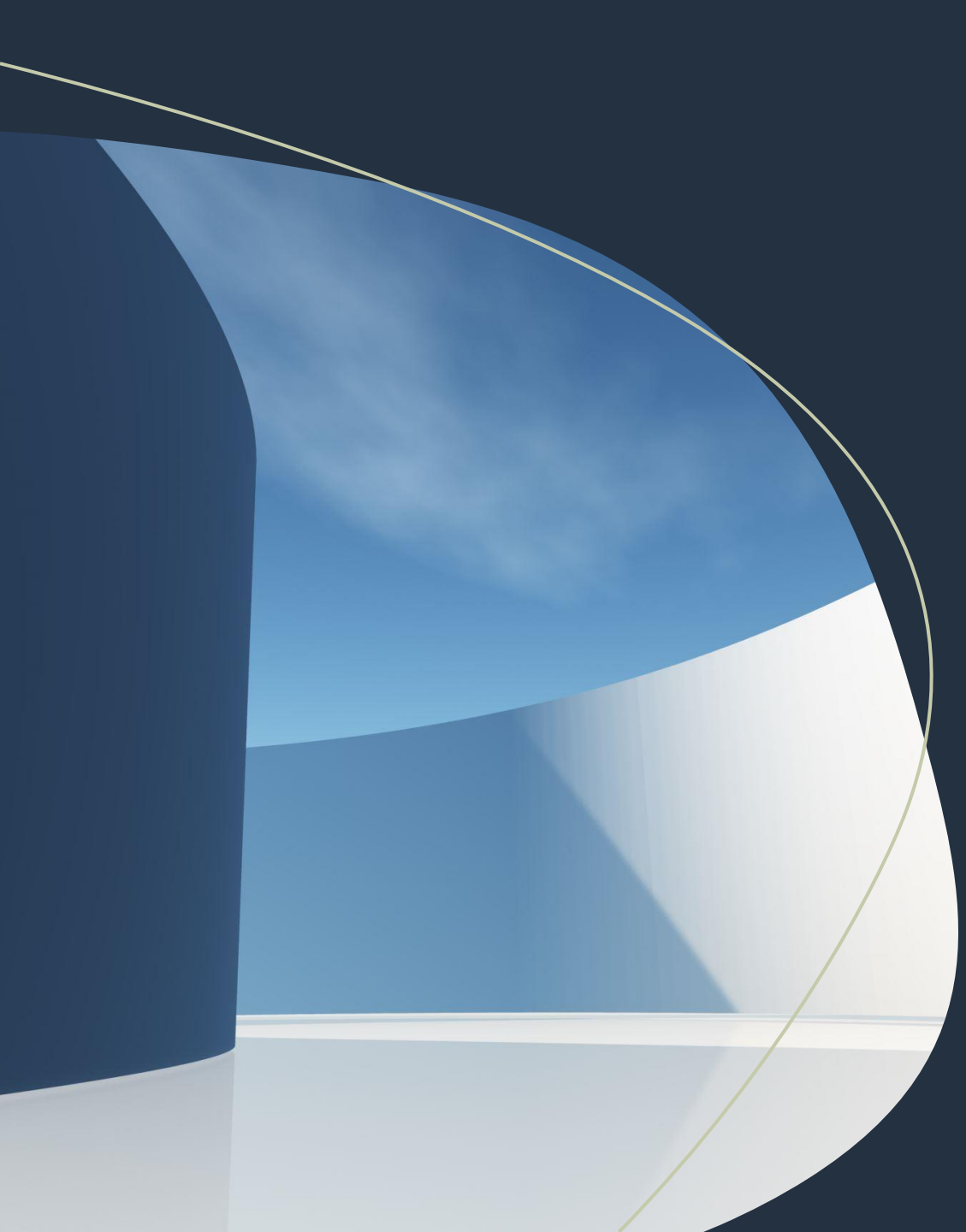


Razpoznavanje števk z LSTM omrežje

Iva Eftimska

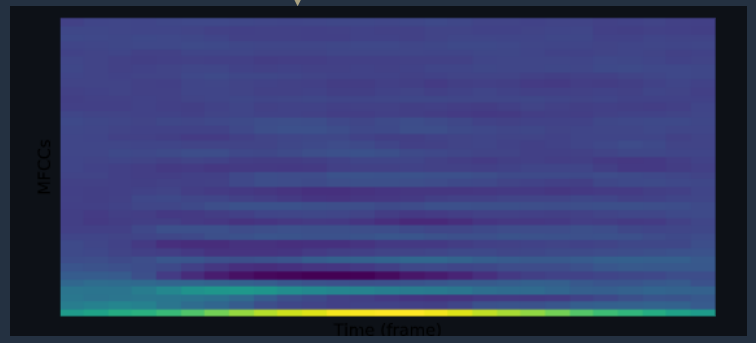
Podatkovna baza

- 3000 posnetkov (50 posnetkov za vsako številko po govorec)
- 6 govorcev
- Različne akcente v angleščini

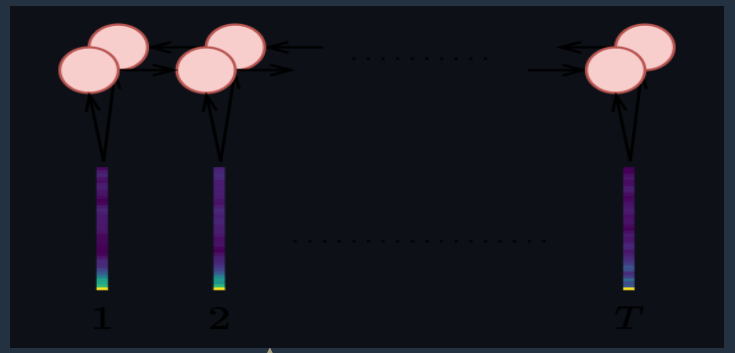
Diagram poteka

MFCC

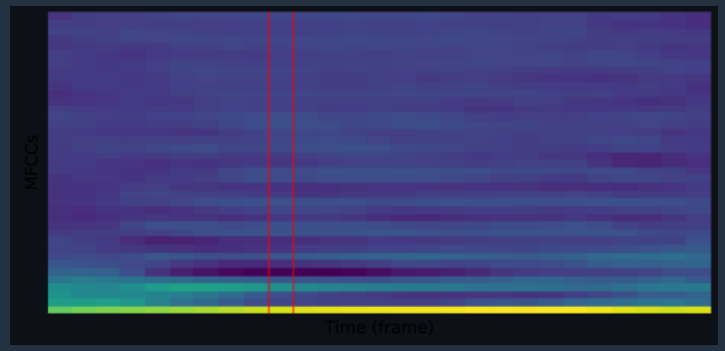
Remove offset coefficent



Per frame MFCC standardization

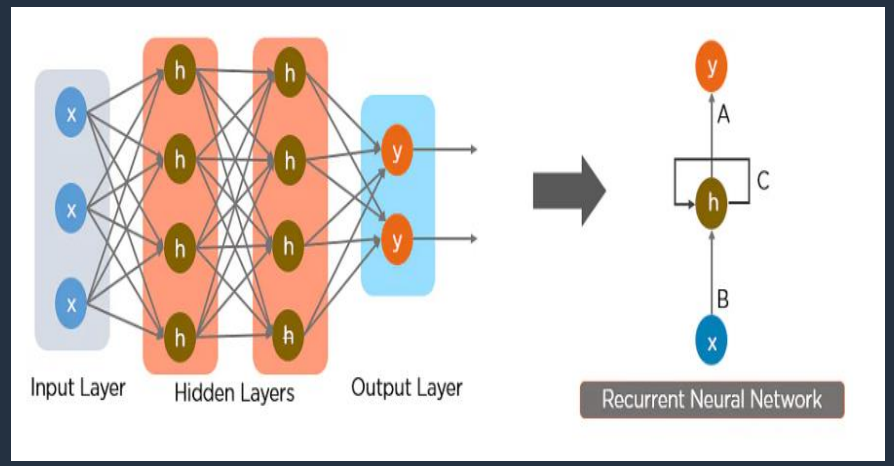


Feed into single layer BiLSTM-RNN

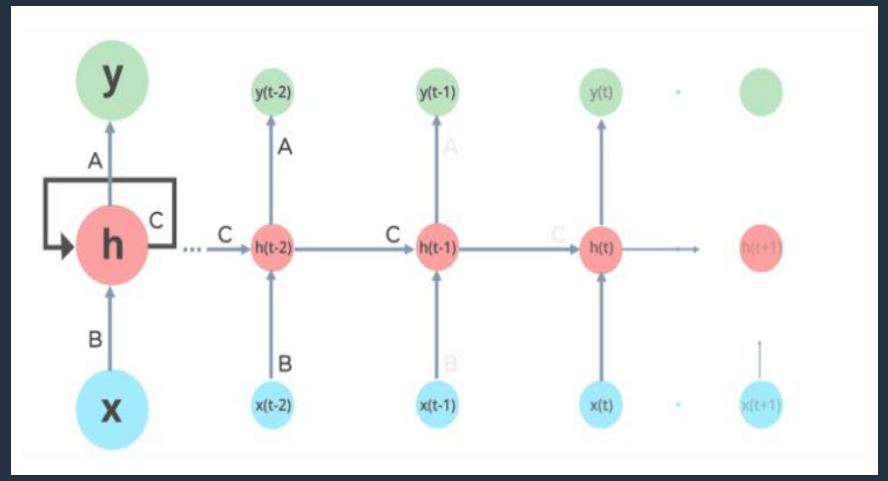


LSTM RNN

- RNN
- Vanishing gradient problem

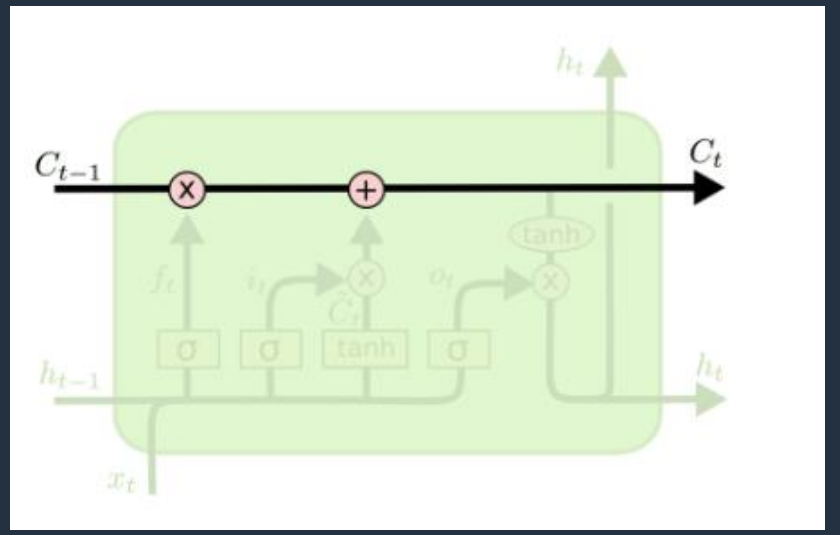


$$\mathbf{h}_{\langle t \rangle} = f\left(\mathbf{h}_{\langle t-1 \rangle}, x_t\right)$$

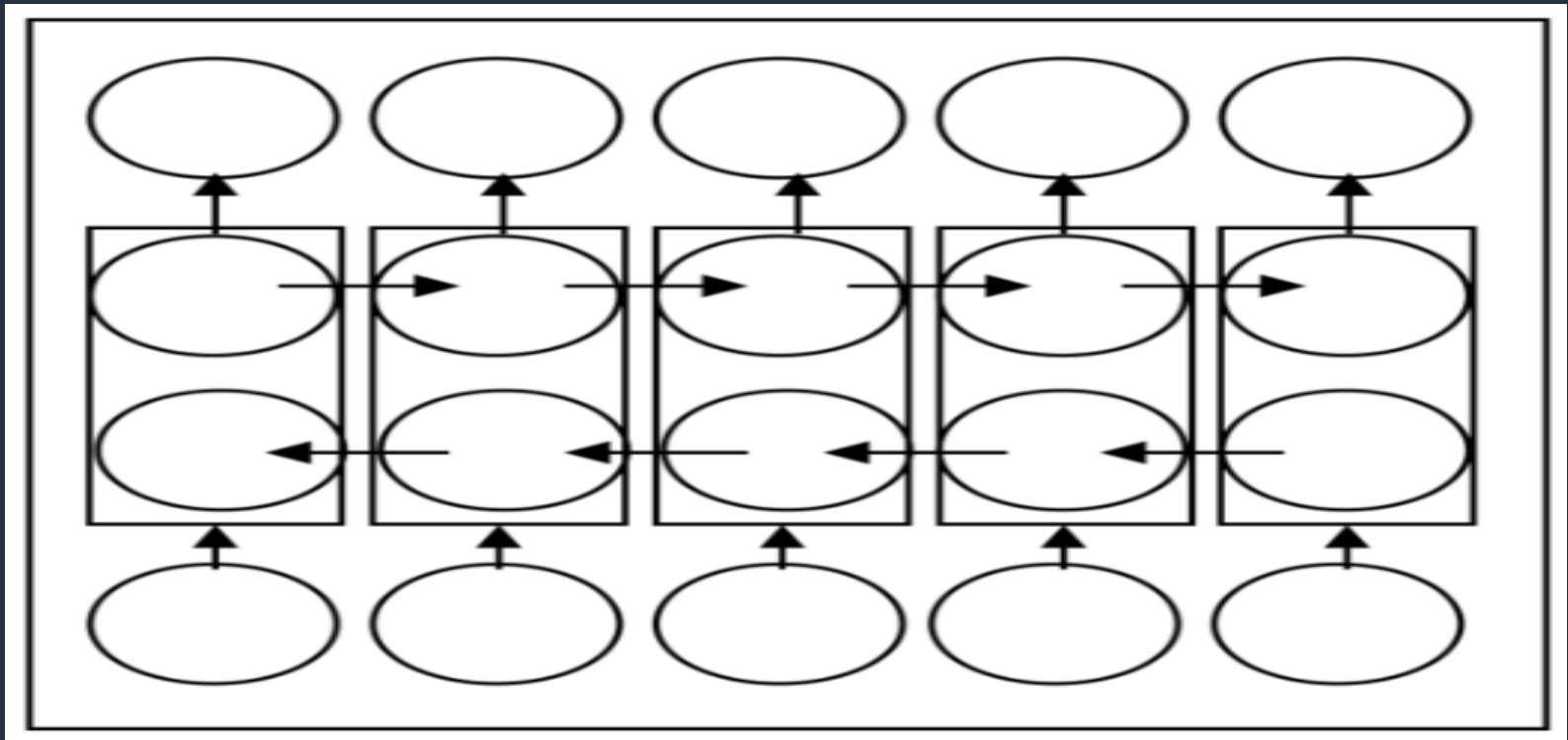


LSTM:

Bolj odporne na vanishing gradient problem



BiLSTM-RNN



Izvedba v Pytorch

- 1) Izločanje MFCC značilk (39) in predobdelava značilk, povprečje 0, std deviacija 1
- 2) Deljenje podatkovne baze, 80% učenje, 10% validacija, 10% testiranje
- 2) Arhitektura modela
- 3) Štimanje parametrov
- 4) Rezultate

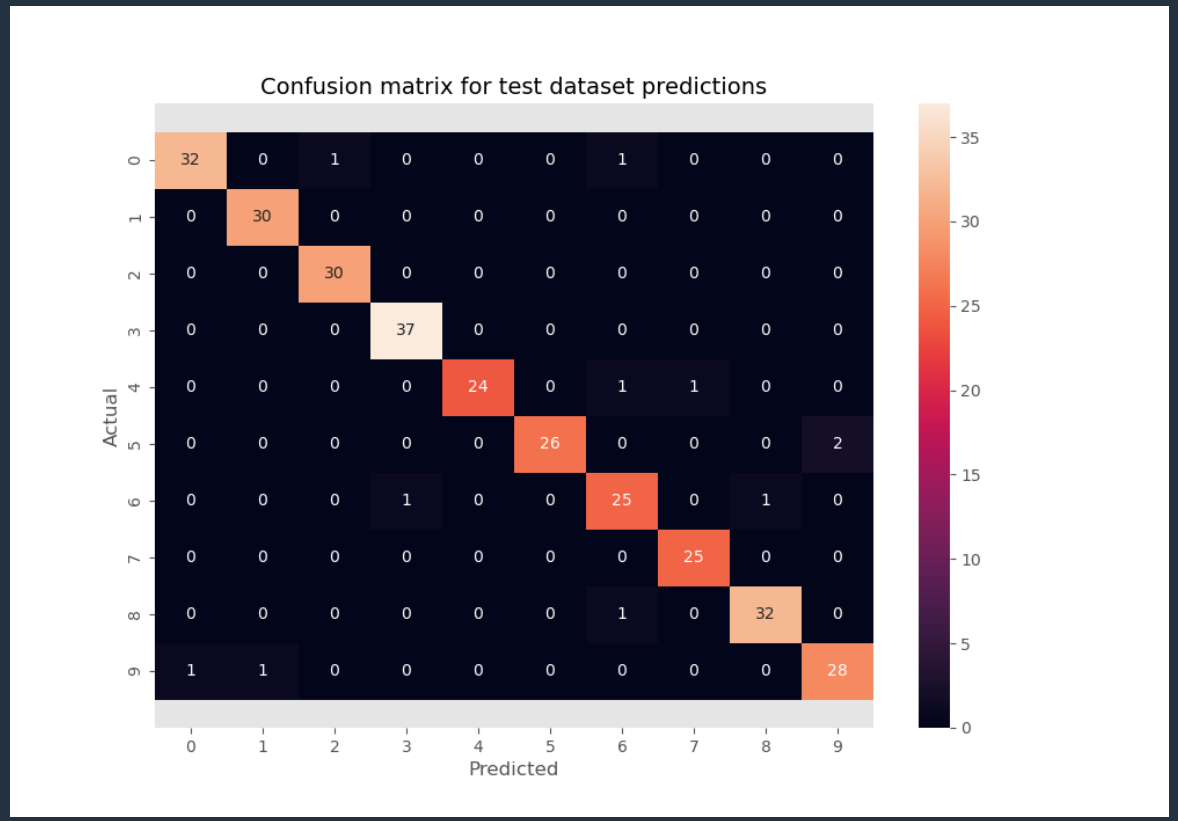
Model v Pytorch

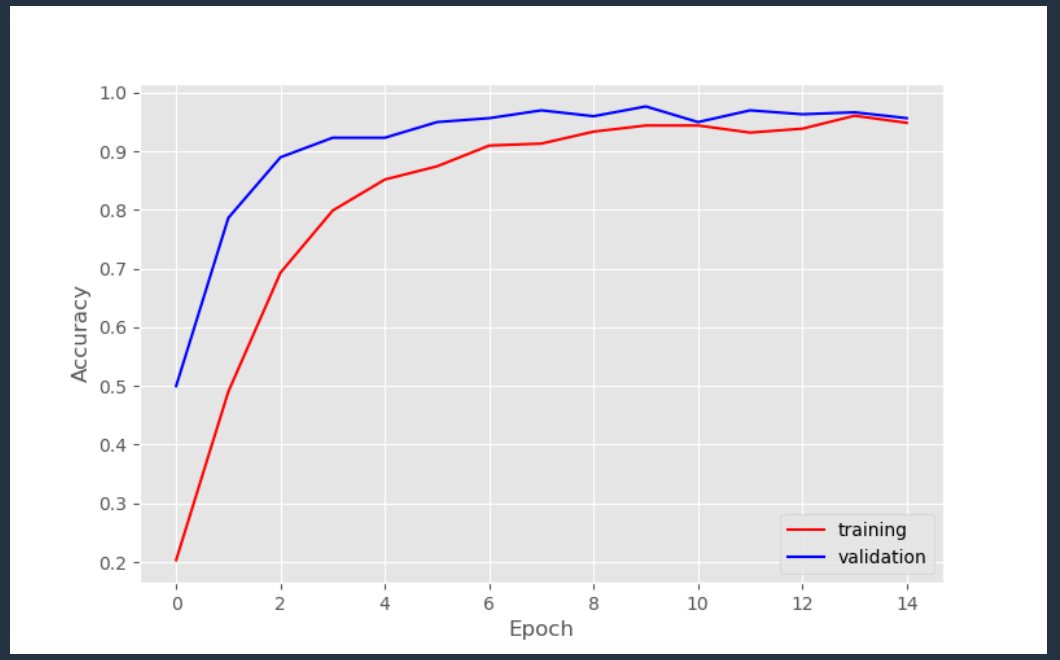
```
def __init__(self, n_mfcc, n_label, h, d, n_lstm):
    super().__init__()
    self.lstm_layer = nn.LSTM(input_size=n_mfcc, hidden_size=h, num_layers=n_lstm, batch_first=True, bidirectional=True)
    self.lstm_layer_dropout = nn.Dropout()
    self.linear_layer = nn.Linear(in_features=h*2, out_features=d)
    self.linear_layer_relu = nn.ReLU()
    self.linear_layer_dropout = nn.Dropout()
    self.output_layer = nn.Linear(in_features=d, out_features=n_label)
    self.output_layer_logsoftmax = nn.LogSoftmax(dim=1)

def forward(self, x, lengths):
    batch_size = len(x)
    x = pack_padded_sequence(x, lengths.to('cpu'), batch_first=True)
    x, (hn, cn) = self.lstm_layer(x)
    hn = self.lstm_layer_dropout(hn)
    hn = hn.transpose(1, 2).reshape(-1, batch_size).transpose(1, 0)
    hn = self.linear_layer_relu(self.linear_layer(hn))
    hn = self.linear_layer_dropout(hn)
    return self.output_layer_logsoftmax(self.output_layer(hn))
```

- Batch size (unchanged) : **64**
- Number of epochs (unchanged) : **15**
- Learning rate (unchanged) : **0.002** with **Adam Optimizer**
- Number of MFCCs : **39**
- Number of LSTM layer : **1**
- Number of Hidden state dimensions : **50**
- Number of units in Linear Feed forward Neural Network : **50**

Rezultate






```
Train Dataset accuracy: 94.88
Validation Dataset accuracy: 95.67
Test Dataset Accuracy: 96.33
modi1s9@evaluks9:~/Desktop/govorne_seminarska/spoken_digi
```