Varying Number of Features of Neural Network Input Objects

1 Problem

In a feed forward neural network, each feature of the input objects corresponds to an input neuron of the neural network. But it is unclear what to do if the number of features of the input objects varies.

In the case of particle physics the input objects can be events in a collider experiment and the purpose of the neural network can be to distinguish between signal and background. Figure 1 shows a part of such a network. In this example, the fourth input neuron takes p_{T} of jet number seven. The question is what to do if there are events with fewer than seven jets.

2 Possible Solution

A possible solution for this problem is an additional input neuron for each jet. The input of these additional neurons is $x_i = 0$ if the corresponding jet i exists and $x_i = 1$ if the corresponding jet i does not exist. If the jet does not exist, the neurons that would
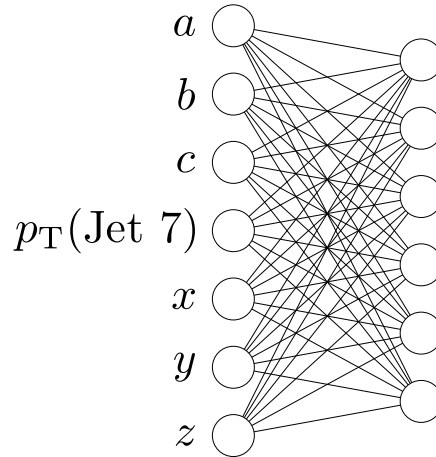
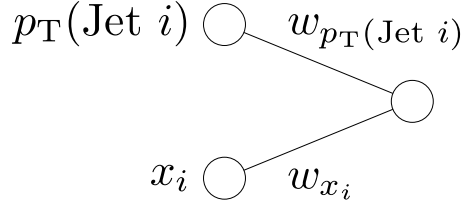

Figure 1: Part of a neural network.

Figure 2: Part of a neural network with additional input neuron.

take the properties of the jet (e.g. p_T) as their input get an arbitrary but fixed value. Equation (1) is a part of the input of the neuron in the first hidden layer (see figure 2).

$$\cdots + p_\mathrm{T}(\text{Jet } i) \cdot w_{p_\mathrm{T}(\text{Jet } i)} + x_i \cdot w_{x_i} + \cdots \tag{1}$$

If jet i exists, the second summand of equation (1) is zero and the input for the hidden neuron is not influenced by the additional input neuron. If jet i does not exist, the weight w_{x_i} can be used to push the result into the direction of the wanted output for this case.

3 Results

The procedure described in section 2 was tested with a network that separates $\mathrm{t\bar{t}H(b\bar{b})}$ signal events from $\mathrm{t\bar{t}}$ + jets (including $\mathrm{t\bar{t}b\bar{b}}$) background events. For comparison, a second network was trained where the jets that do not exist for all events were not used as input at all. The network with the additional jets did not perform better than the other one.

4 Open Questions

- Is there a case where the procedure described in section 2 improves the performance of a neural network?

- If there is such a case:

  - Is it important, that the additional variable is $x_i = 0$ if the corresponding jet i exists and $x_i = 1$ if the corresponding jet i does not exist? Or can you also use the usual system (which is the other way round)?

  - Is it a problem to include the additional variables in the feature scaling?

  - Should the feature scaling happen before or after the not existing variables are set to a fixed value?

  - A possible alternative to the procedure described in section 2 is to introduce more jet btag categories (with 7 jets, 8 jets, …). This is a simpler procedure, but there are fewer training events per category available. Which solution performs better?

5   Code Fragment

The code fragment in the file `preprocessing_fragment.py` shows how the procedure described in section 2 can be implemented using the Pandas DataFrame.