be just as useful for fastai v2. There is no 2020 version of part 2 of the course recorded yet, and we don't have a date for when that might happen.

The 2020 version of the course includes material covering both machine learning *and* deep learning. So there won't be a separate "Introduction to Machine Learning" (although the old one will still be available).

fastai v1 will continue to be available, and we'll continue to provide bug fixes (and accept pull requests for it). To *pin* your fastai version to v1 (i.e., to avoid it upgrading automatically to v2), run the following command (assuming you use conda):

```
echo 'fastai 1.*' >> $CONDA_PREFIX/conda-meta/pinned
```

Then, when you're ready to upgrade to v2, remove the $CONDA_PREFIX/conda-meta/pinned file.

Order of layers in model

ankitbko Ankit Sinha   Feb '17

In general when I am creating a model, what should be the order in which Convolution Layer, Batch Normalization, Max Pooling and Dropout occur?

Is the following order correct -

```
x = Convolution1D(64, 5, activation='relu')(inp)
x = MaxPooling1D()(x)
x = Dropout(0.2)(x)
x = BatchNormalization()(x)
```

In some **places** I read that Batch Norm should be put after convolution but before Activation. Even **ResNet** has similar structure. Something like this -

```
x = Convolution1D(64, 5)(x)
x = BatchNormalization()(x)
x = Activation('relu')(x)
x = MaxPooling1D()(x)
x = Dropout(0.2)(x)
```

However in **Lesson 7** Batch Normalization happens after Conv + Activation but before Max Pooling.

In Lesson 3, when Batch Normalization was first introduced, though it was used in FC layers, it was placed after Dropout.

Does the ordering of these layers matter? Which order is considered to give best result?

geniusgeek Samuel Ekpe   Feb '17

First you have to Understand what each of them does, and understand the layer/model you are trying to build and the approach being taken, the structure matters.

Max pooling :is a sample-based discretization process. The objective is to down-sample an input representation (image, hidden-layer output matrix, etc.), reducing it's dimensionality and allowing for assumptions to be made about features contained in the sub-regions binned.

Batch Normalization: normalization (shifting inputs to zero-mean and unit variance) is often used as a pre-processing step to make the data comparable across features. This therefore leads to higher learning rate and better speed.

To learn more about Batch Normalization, take a look at:
https://www.quora.com/Why-does-batch-normalization-help

Drop out: Its a regularization functino that reduces/prevents overfitting, by normalizing some pixels to 0.

Dilution (neural networks)

Dilution (also called Dropout) is a regularization technique for reducing overfitting in artificial neural networks by preventing complex co-adaptations on training data. It is an efficient way of performing model averaging with neural networks. The term dilution refers to the thinning of the weights. The term dropout refers to randomly "dropping out", or omitting, units (both hidden and visible) during the training process of a neural network. Both the thinning of weights and drop Dilution is us...

JMLRdropout.pdf
2.75 MB

Also regarding the values in the CNN layer, these correspond to the dimension, size and depth of the image. So when you are adding another layer, you must take into consideration to the output of the previous layer(ie output dimension of MaxPooling) as input for the new layer.

For the FC layer it squashes the inputs to the expected output using the defined output dimension was 4096(64X64).

The long and short is that you need to know what you are doing!, read papers, take a look at how models are built in such layers. feel free to ask questions when stuck!

---

drscotthawley Scott H Hawley   **Feb '20**

I realize this is an old thread, but given that it appears near the top of Google results on such a topic, and that the above reply doesn't even attempt to answer the question of ordering, I want to leave this here:

An important point is that monotonic activation functions commute with (max- or average-)pooling. This means that the order *does not matter*. So you might as well save some time and do the pooling first, thereby reducing the number of operations performed by the activation.

Same thing goes for batch norm…to an extent: Whether you put it before or after your activation is a matter of some opinion, but putting it before or after MaxPooling will make very little difference on the accuracy – yet will affect the speed.

Similarly for Dropout: it commutes with many activations such as ReLU and tanh – any function f for which f(0)=0 – so the order doesn't matter. Doing Dropout before or after BN will make a small difference but for large layers (or not-too-much dropout) the different will be negligible. For large dropouts & small number of

neurons,… you'll see some variability on the ordering. Dropout before or after pooling? As you noted, usually it appears after pooling.

This commutivity (commutativity?) property is one reason why you'll sometimes see layers ordered differently: because it may not affect the results. But it *can* affect execution time! 😉

Note that BN and ReLU do *not* commute, and people's choices seem to vary on which they do first. For more on that, see Sylvain's reply on this related thread: **Where should I place the batch normalization layer(s)?** , where he notes that that FastAI default is to follow ResNet and do BN before ReLU.

But other authors will do differently. For example, in this post on **why the idea that BN cures internal covariant shift is a myth** , it's noted that "it has been found in practice that applying batch norm after the activation yields better results." For *them*, for *their* problem. Try reversing the order on your problem, and use whatever works best.

What about non-monotonic activations like Mish? I haven't tried. Mish is still *close* to monotonic for most inputs, it just has that little "dip" to the left of zero, which will affect some results. My intution suggests that you could still put it after pooling and save time, but check with those who do this.

EDIT: By the way, these and other things I learned from a great post that Jeremy once shared: **https://myrtle.ai/how-to-train-your-resnet-8-bag-of-tricks/**

Feb 2017

2 / 3
Feb 2017

Feb 2020