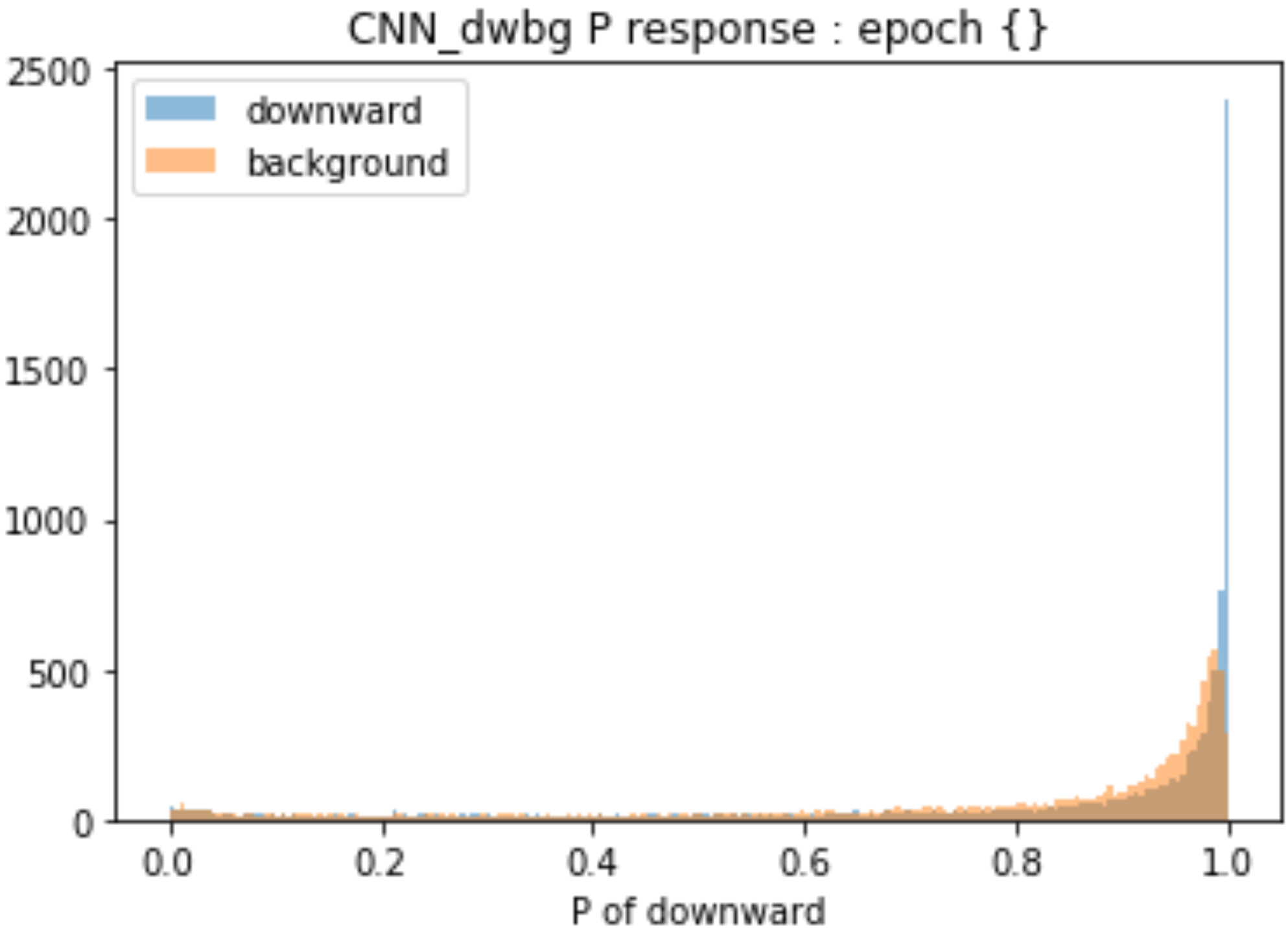
Deep Learning in GBAR

Undergraduate Intern
Byeongyoon Park

Why Training and Validation results are different?

```
Epoch 13/50
180000/180000 [==============================] - 103s 571us/step - loss: 0.1341 - categorical_accuracy: 0.9465 - acc_dw: 0.91
24 - acc_bg: 0.9811 - val_loss: 27.7825 - val_categorical_accuracy: 0.5000 - val_acc_dw: 1.0000 - val_acc_bg: 0.0000e+00
Epoch 14/50
```

Response Histogram



Overfitting

- Definition (By Wikipedia)

In statistics, **overfitting** is "the production of an analysis that corresponds too closely or exactly to a particular set of data, and may therefore fail to fit additional data or predict future observations reliably".

An **overfitted model** is a statistical model that contains more parameters than can be justified by the data.

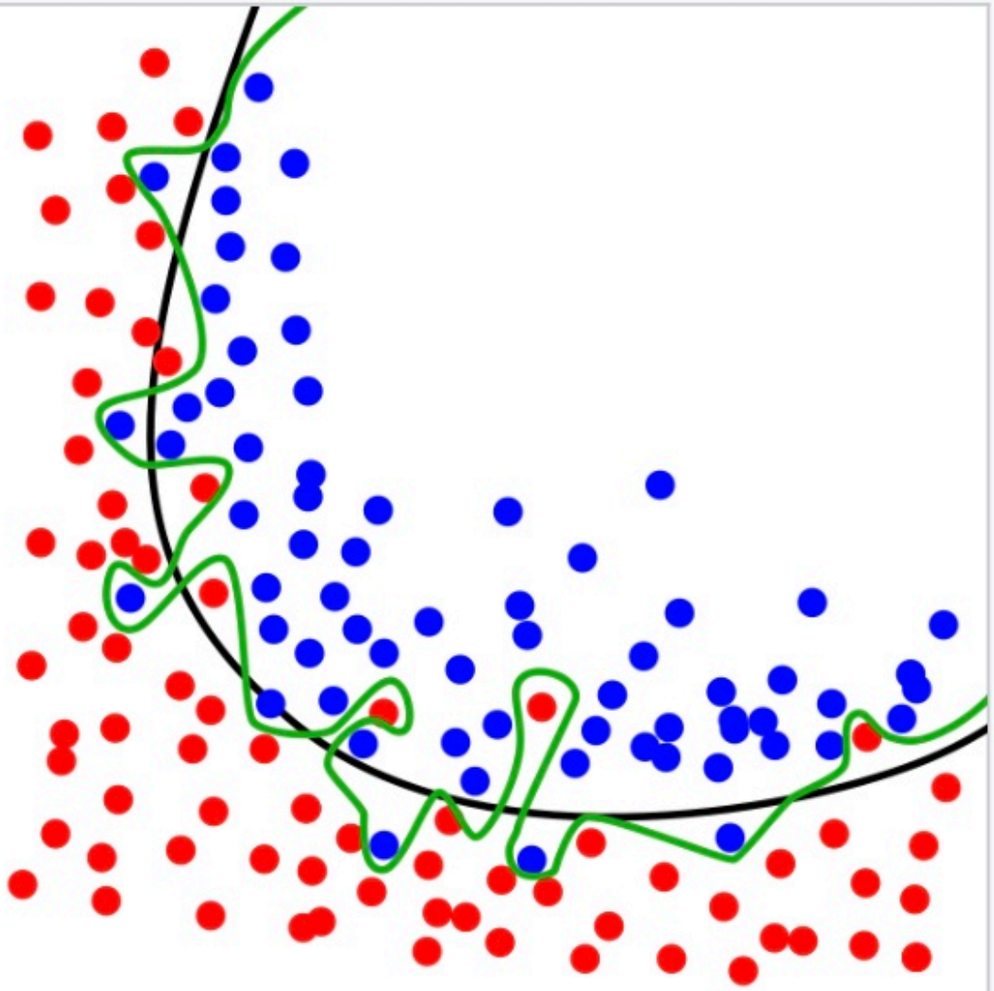


Figure 1. The green line represents an overfitted model and the black line represents a regularized model. While the green line best follows the training data, it is too dependent on that data and it is likely to have a higher error rate on new unseen data, compared to the black line.

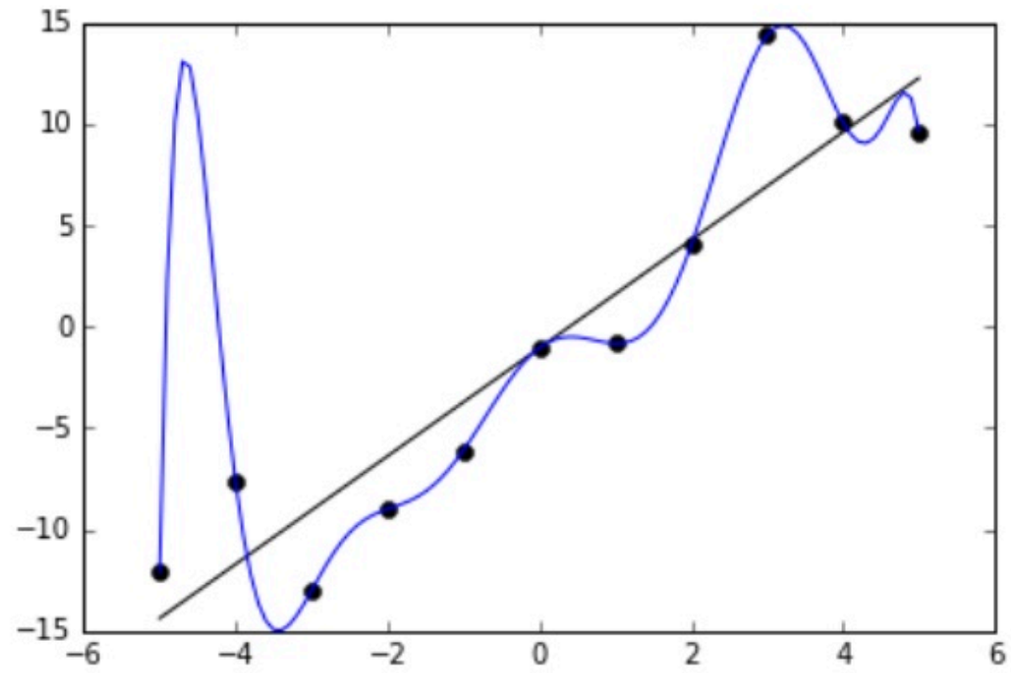


Figure 2. Noisy (roughly linear) data is fitted to a linear function and a polynomial function. Although the polynomial function is a perfect fit, the linear function can be expected to generalize better: if the two functions were used to extrapolate beyond the fitted data, the linear function should make better predictions.

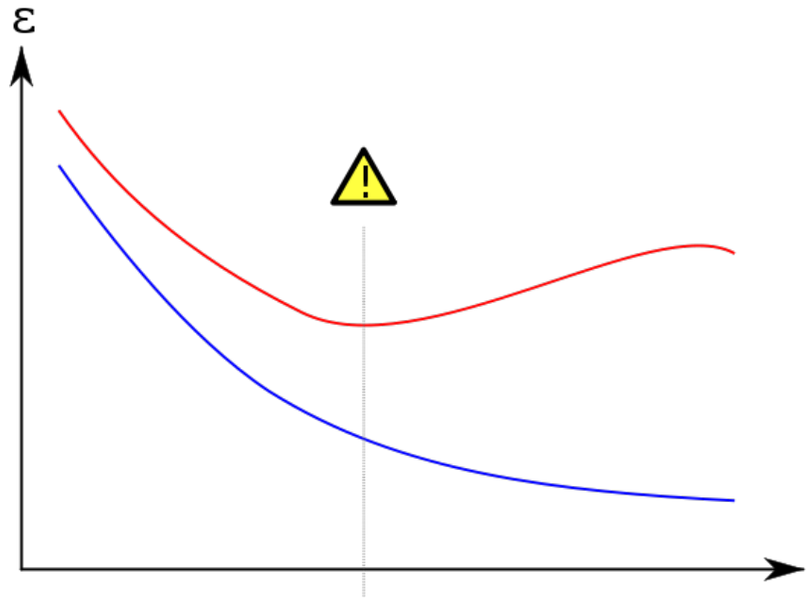
Wikipedia, Overfitting

Overfitting in ML

- Occam's Razor

"Entities should not be multiplied without necessity."

'... and the complex overfitted function will likely function will likely perform worse than the simpler function on validation data outside the training dataset, even though the complex function performed as well, or perhaps even better, on the training dataset.'



Why Overfitted? - Possibilities

1. There is a bias in the training dataset than that of test.

2. The architecture of current CNN is so complex that the overfitting easily occurs.

3. The internal ‘functions’ are not appropriate for our signal waveform sequence, for image recognition(MNIST).

Solutions

1. We will use the dataset of more than 2 hits.

2. 3. We will research for appropriate architecture and internal structure(functions) of CNN to apply on our dataset.