

Vowpal Wabbit

The purpose of these notes is to explain how Vowpal Wabbit works. You will need to apply these concepts to complete your next programming assignment and for the last midterm.

1 Concepts

What is *one-hot encoding*?

What is *the hashing trick*?

Reference: <https://booking.ai/dont-be-tricked-by-the-hashing-trick-192a6aae3087>

What is *progressive online validation*?

How does Vowpal Wabbit use a *validation set*?

2 Important Hyperparameters (command line options)

You will need to tune (some of) the options below in order to achieve good performance on your assignment. You can find more information about each of these hyperparameters either on the web-based documentation at

https://vowpalwabbit.org/docs/vowpal_wabbit/python/latest/command_line_args.html

or by running the terminal command

```
$ man vw
```

2.1 Hypothesis Class / VC dimension

The following options modify the hypothesis class that you are trying to learn, and therefore affect your training and generalization error. These are the most important parameters to tune, and you should begin your assignment by tuning these parameters.

```
--bit_precision
```

```
--hash_seed
```

```
--quadratic ::
```

`--cubic ::`

`--nn`

`--ngrams`

`--skips`

`--feature_limit`

--11

--12

--no-bias-regularization

2.2 Optimization

The following parameters control the optimization process of your model, and therefore control the optimization error. I recommend adjusting these parameters only after you are happy with your hypothesis class.

`--loss_function`

`--binary`

`--passes`

`--learning_rate`

`--random-weights`

`--early-terminate`

`--holdout_off`

3 Problems

Problem 1. You have trained a vowpal wabbit model with the hyperparameters:

```
--bit_precision=22  
--l2=1e-3  
--quadratic ::  
--learning_rate=0.01
```

1. What is the VC dimension of your hypothesis class?
2. If you remove the `--quadratic ::` hyperparameter and keep `--bit_precision` constant, how would you expect to change the `--l2` hyperparameter in order to achieve similar generalization error? (That is, should you increase it, decrease it, or keep it the same?) Why?

Problem 2. You have found that the optimal hyperparameters for a problem you are working on is

```
--bit_precision=12
--l1=1e-3
--passes=5
```

If you acquire significantly more training data, how should you expect to adjust each of these hyperparameters? (That is, should you increase them, decrease them, or keep them the same?) Why?

Problem 3. You have found that the optimal hyperparameters for a problem you are working on is

```
--bit_precision=28
--ngrams=2
--passes=20
--learning_rate=0.01
```

The average loss reported by `vw` when training using these hyperparameters is

```
average loss = 0.045916 h
```

If you add the command line flag `--holdout_off` when training the model, would you expect the average loss reported by `vw` to increase, decrease, or stay the same?