

Production machine learning

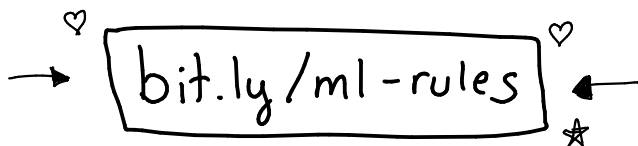


by Julia Evans
twitter: @b0rk
blog: jvns.ca

What's this?

I read "Rules of Machine Learning: Best Practices for ML Engineering" and a lot of it really resonated with my experience doing machine learning.

You can find that PDF at:

→  bit.ly/ml-rules

In this zine, I'll talk about what I've learned about machine learning! Basically I want to talk about 5 of my favourite rules from that PDF.

(there are 43 rules, it's great advice, you should read them all)

Rule #3: Use machine learning instead of complicated heuristics

simple heuristics can be
≡ awesome ≡



hey, look, if I
just recommend
Minecraft to every
8 year old, it works
great

ship it!

but complicated heuristics
are a nightmare

```
if (age > 7 and gender = M  
and (country = "Canada" or  
country = "USA") and  
favorite_color = blue) ...
```



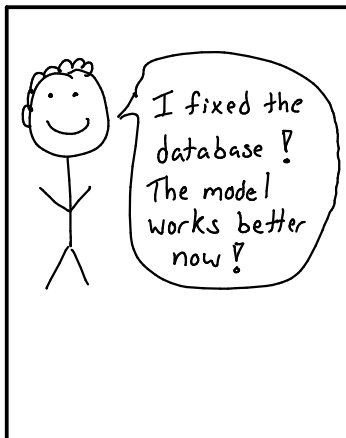
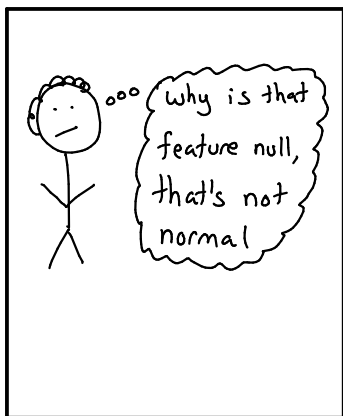
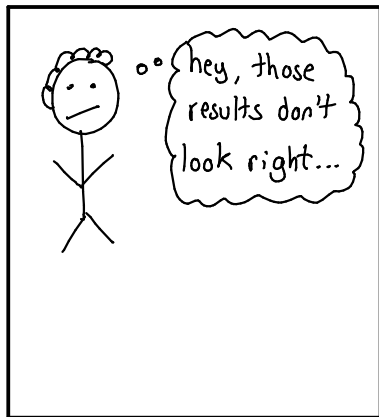
I have regrets

machine learning is much more maintainable
than that mess, and it'll perform better



this model is still a bit complicated,
but at least it's not based on
one ad hoc data analysis
I did 2 years ago !!

Rule # 10: watch for silent failures



Rule #16 : keep your pipeline healthy

this is sad :



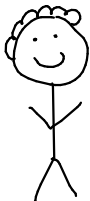
I can't generate the features I need to train a new model !

and this



I'm worried there's a bug + I don't trust our trained models

what we want :

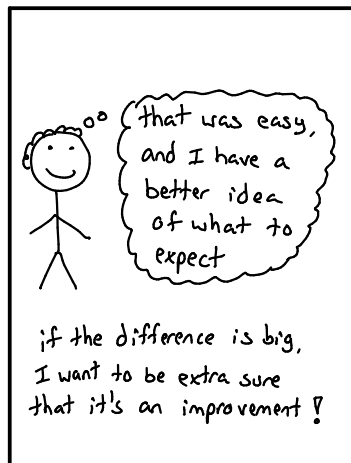
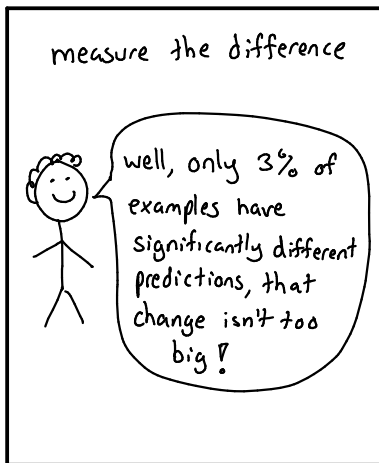
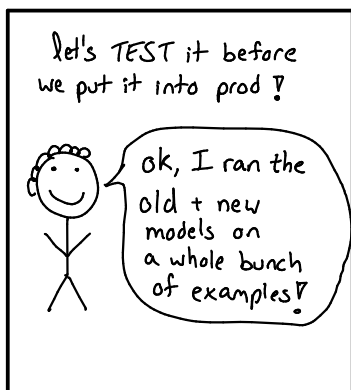
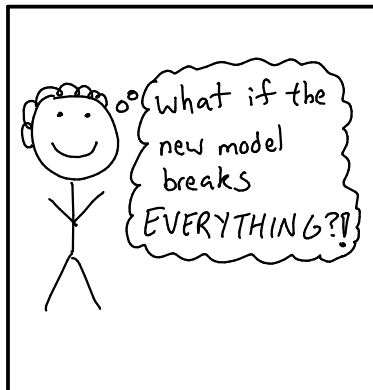


when I have new ideas / data, I can easily launch a new model !

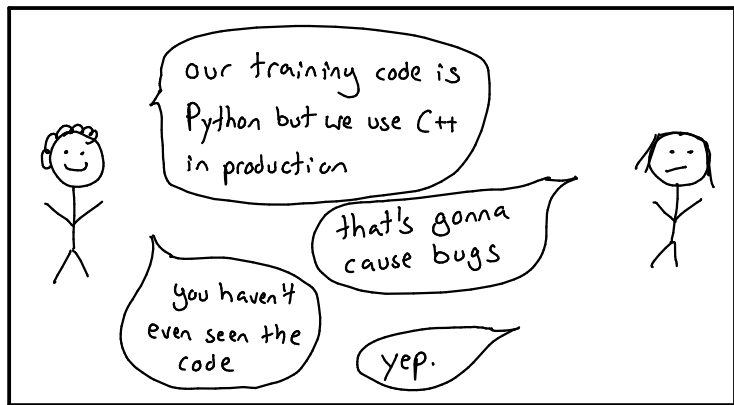


we're careful about adding complexity !

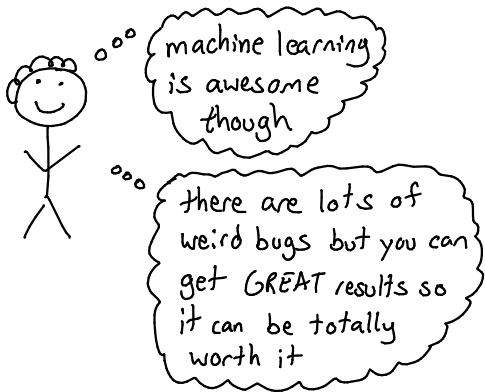
Rule #24: Measure the change between models



Rule # 32 : Reuse code between training + production when you can



- you're gonna have bugs related to training / production differences
- you probably can't make this perfect
- just try to monitor differences.



♡ this? There are more
of these at

→ jvns.ca/zines ←