9-6-18

meeting w/ Bill Noble, Jacob, Ariel

ROC area = accuracy of putting a positive above
a negative

spectrum kernel → count aa, pairs, triples, etc.
- sparse data = ok
- triple data can get aa back
  to original seq (30-mer)
- see how ea information
  does in combination w/ others.

→ regularize model to account for
Sparseness — Nobel has way to do this.

→ use multiple values of k simultaneously

_____

can sort by rank instead of
activator score — a way to avoid
exponential data.

①

Ariel: Dense network = significant
improvement

$\qquad$ neurons → got

↳ does small #
neurons mean a
consice signature?

correlate features w/ output

~~Ariel~~ bias for negative charges in
AD positives — but this is not the
only important feature

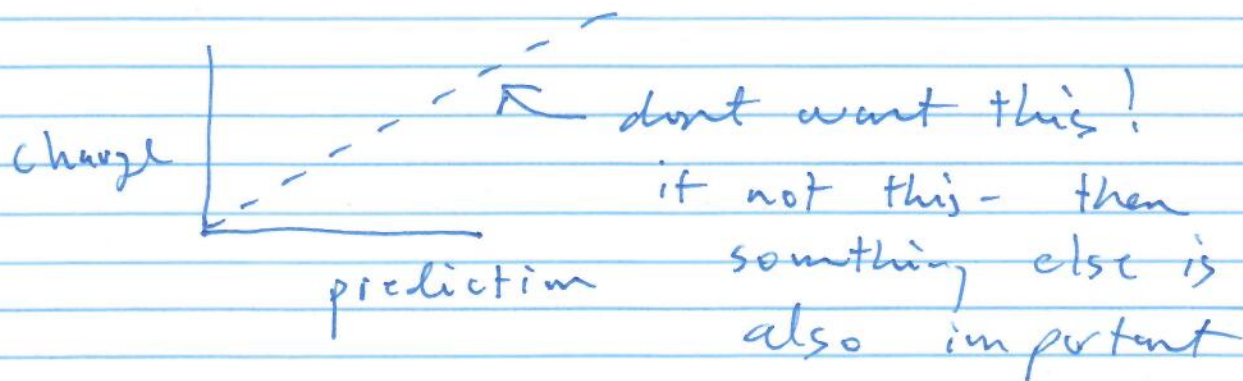make sure that we do better
than just looking at charge.

Accuracy # matters depending on system
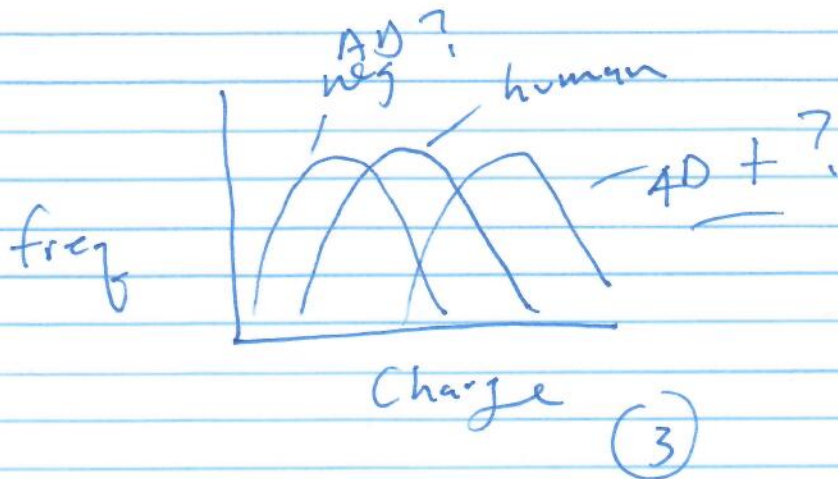
Stock market -94% accuracy $= \$\$$

self-driving car -94% $= \widehat{x\ x}$

need to do better than charge alone

try:

charge

prediction

← dont want this!
if not this — then
something else is
also important
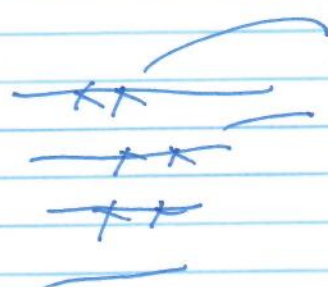
try w/ random sequences from human genome (30-mers?)

freq

AD?
neg   human
                −AD +?

Charge

(3)

what is hot-encoding?

will not reveal motif.

~~**x x**~~ what are important features?

~~**x x**~~

~~**x x**~~ which lead to AD$^+$?

Jacob supposedly has package that
can reveal most important features
in every seq.

more fine-grained info

Ariel will try to implement these
suggestions & get back to Jacob/Bill

(4)