# theory, data, and model selection

## intro

### caveats

( “so let me start with warnings and disclaimers”)

1. warning: i am not a population geneticist
2. disclaimer: i will likely not talk about **any** of my own work.
3. “informal” talk, please ask questions

## statement of the problem:

1. quest for ``truthiness” / using data to select best model
2. examples
   I. comparing models: neutral, neutral+mutation, neutral+selection, neutral+migration+sweeping
   II. comparing # features: 50 sites, N-50 hitchhikers, 100 sites, N-100 hitchhikers, N sites 0 hitchhikers?
3. key ideas
   I. x^2 ergo ML ergo ME (jets)
   II. prediction ergo CV (sharks)
4. when would this be a bad idea? example: NN-regression
   I. good models are predictive [but/and]
   II. good models are interpretable

## basic bayesian notions

1. context + historical digression: who was this bayes [http://upload.wikimedia.org/wikipedia/en/d/d4/Thomas_Bayes.gif] troublemaker?
2. what bayes said: “the product rule”
3. (as opposed to “the sum rule”)
4. examples of what happens when you put product and sum together
   I. diffusion
   II. likelihood with additive and normal noise
   III. behold: why we fit

## bayesian model selection

1. if you believe in ML, why not MML=ME?

I. plot of P(D|K)
II. being bayesian: parameters and priors
III. payoff: BIC
    a. WARNING: SKETCH OF DERIVATION LIKELY [http://www.columbia.edu/itc/applied/wiggins/Movies/BIC.tiff]
    b. “the razor”
    c. “truthiness”
    d. summary + swindles
    e. elaboration on priors
        α. gaussian priors (exact case for linear regression; restatement of ridge/Tychonoff regression)
        β. other priors: lasso, grouping, fusion, etc.

2. who is the most bayesian? bayes’ rule, etc.
3. a note on graphical models

## frequentist ideas / cross validation

1. modeling is about prediction, ergo minimize empirical estimate of generalization error.
2. fit (monotonic) & “truthiness” (peaked) plot
3. nature never hands you distributions, only observations (re: P(D|M), BIC)

## illustration: mixture modeling

1. cf xing
2. cf sharply peaked
3. connection w/stat mech + test distributions + mean field theory + BP

## bayesian+variational

1. “EM” movie [http://www.columbia.edu/itc/applied/wiggins/Movies/jmh_gmm_demo.mov]
    I. WARNING: SKETCH OF DERIVATION LIKELY [http://www.columbia.edu/itc/applied/wiggins/Movies/VB_examples.pdf]
    a. gibbs=jensen=feynman inequality
    b. ln q(x) \propto < ln p(D,z,t|a,b) > _{ ~x } = dH/dq
    c. q=exp(dH/dq)
    d. < H >_{H’} = H’
        α. recall, as you learned on your mother’s knee: m = tanh ( Jm/T)
        β. cf section 33.3 of mackay’s book
2. “VB”/“ensemble learning”
    I. E: q(z)
    II. M: q(t)
    III. Movie of VB in action [http://www.columbia.edu/itc/applied/wiggins/Movies/flyingGaussians1.avi] for vector GMM

3. a note on graphical models
   I. gmm ML graphical model, c/o jake hofman [http://www.columbia.edu/itc/applied/wiggins/Movies/jmh_gmm_ml.pdf]
   II. gmm ME graphical model, c/o jake hofman [http://www.columbia.edu/itc/applied/wiggins/Movies/jmh_gmm_me.pdf]

## CV approach

(it works. discuss)

# pvalueology

1. how good is my CV?
2. how good is my likelihood
3. assumptions inherent
4. conventions inherent
5. multiple hypothesis testing

# alternative worldviews not mentioned in this talk

1. discriminative learning, SVM/hinge loss, boosting loss...
2. model selection via stability
3. sampling/MCMC/gibbs (vs variation)

# topics for penalty time

1. “data mining”
2. work on community detection as latent variable inference (inc. Hofman+Wiggins '08 [http://arxiv.org/abs/0709.3512])
3. deep thoughts: what is a model? what is a “good” model?

# references

## statistics books by physicists

1. mckay
2. bishop

## online references

1. beal's thesis on variational methods [http://www.cse.buffalo.edu/faculty/mbeal/thesis/]
2. yedidia's lecture notes [http://www.merl.com/people/yedidia/santafe.pdf]

3. wikipedia pages
   I. resampling statistics [http://en.wikipedia.org/wiki/Resampling_(statistics)]
   II. Gibbs sampling [http://en.wikipedia.org/wiki/Gibbs_sampling]
   III. Variational Bayes [http://en.wikipedia.org/wiki/Variational_Bayes]
   IV. Bayes [http://en.wikipedia.org/wiki/Thomas_Bayes]
   V. Ridge/Tichinoff/Bayesian regression [http://en.wikipedia.org/wiki/Ridge_regression]

## misc papers

1. beal+ZG: vb vs other things [http://learning.eng.cam.ac.uk/zoubin/papers/valencia02.pdf]
2. more yedidia [http://www.merl.com/reports/docs/TR2000-27.pdf], on corrections to MFT
3. still more yedidia [http://nerdwisdom.files.wordpress.com/2007/10/ja910924.pdf], diagrammatica, less pedagogical.
4. mackay 1995: Probable networks and plausible predictions-a review of practical ... [http://www.inference.phy.cam.ac.uk/mackay/network.pdf], p(D|M) figure; truthiness plots, the whole shebang. 41 pages.
5. Schwarz 1978 [http://www.math.tau.ac.il/~yekutiel/MA%20seminar/Schwarz%201978.pdf]

## deep thoughts

1. share your code
2. "Everything should be made as simple as possible, but no simpler." [http://en.wikiquote.org/wiki/Einstein]
3. "There is always an easy solution to every problem ‚Äî neat, plausible and wrong" [http://en.wikiquote.org/wiki/H._L._Mencken]
4. With four parameters I can fit an elephant, and with five I can make him wiggle his trunk. [http://en.wikiquote.org/wiki/John_von_Neumann]

/Users/wiggins/Sites/dokuwiki/data/pages/kitp_public.txt · Last modified: 2008/10/02 10:10 by wiggins