Neuro 316Q6 — Session 1 1/26/2022

Intro & Course Background

- "modular" structure of probabilistic models
  - ↳ plug & play pieces
  - framework for thinking about prob. modeling

- this course is just a starting point → introduce the main components

- TODO b/w sessions:
  1. reading book & paper(s) ← make sure to use "Perusal" webapp.
  2. exercises & "breif write-up"
  3. complete quiz by noon of day of class (diagnostic)

Modeling

- build a model for $p(y|x)$
  - ↳ "statistical encoding model"

- latent variable model (lower dimensional latent representation)
  - w/ unknown variable z believed to exist: $p(y|z)p(z)$
  - can extend to time: $p(y_t|z_t)p(z_t|z_{t-1})$ "latent dynamics"

Model desiderata:

fitability/tractability ⟷ richness/flexibility

ex.: "linear Gaussian" "GLMs" "neural state space models" "multi-compartment" ← IDK ...

goal of this course

not generally feasible

Bayesian recap

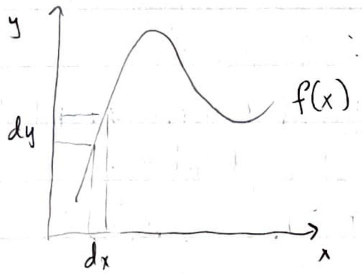
1. prob & prob. models
2. inference of prob. models
3. model comparison

1. Probability & Probabilistic Models

- stimulus x & response y
- want a parametric distribution to describe $p(y|x)$

	$p(y\|x)$	parameters
Poisson	$\text{Poisson}(y\|\lambda(x))$	$\lambda(x)$
Gaussian	$\mathcal{N}(y\|\mu(x), \sigma^2(x))$	$\mu(x), \sigma^2(x)$

- prob dist: a fxn that maps from x to a probability
  - $p(X=x) \equiv p(x)$ is the prob. that r.v. X is x
  - for a PDF, $p(x) \in [0, \infty]$ } ★ important distinction
    PMF, $p(x) \in [0, 1]$
- transforming b/w two r.v.:



$$p(x)\,dx = p(y)\,dy$$
$$p(y) = p(x)\left|\frac{dx}{dy}\right|$$
$$p(y) = p(x)\left|\frac{1}{f'(x)}\right|$$

Probabilities

		x	
		0	1
y	0	0.1	0.2
	1	0.3	0.4

marginal across y

e.g. $p(x) = p(x|y=0) + p(x|y=1)$

e.g. $p(y|x=1) = \overset{\text{joint}}{p(y, x=1)} / \overset{\text{marginal}}{p(x=1)}$

independence: ex: $p(y|x) = p(y)$ then y is indep. of x

x & y are independent if: $p(x,y) = p(x)p(y)$ ← "product rule"

- Bayes rule : $\dfrac{\overbrace{p(y|x)}^{\text{likelihood}}\,\overbrace{p(x)}^{\text{prior}}}{\underbrace{p(y)}_{\text{normalize to } \Sigma \to 1;\ \text{marginal likelihood}}}$

  ↳ basic consequence of standard rules of prob.

- Independent & Identically distributed data (i.i.d data)
  - one data point doesn't affect another
  - data come from the same generative dist

2. Inference w/ probabilistic models

- max. likelihood estimate (MLE)
  - find parameters that make the observed data most likely
  - maximizes the likelihood
  - pros :

pros :	cons :
consistent & converge	noisey for little data
efficient to compute	no estimate of uncertainty

- Bayesian inference
  - most likely model params. : $p(\overset{\theta}{\text{params.}} \mid \overset{y}{\text{data}})$

$$p(\theta | y) = \frac{p(y|\theta)\,p(\theta)}{p(y)} \propto p(y|\theta)\,p(\theta)$$

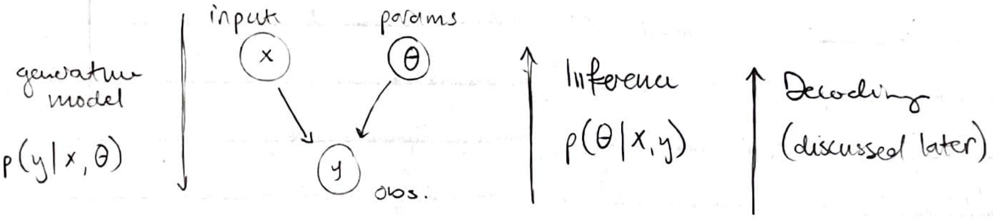
$p(\theta)$: prior beliff of prob. of params θ

$p(y|\theta)$: likelihood of data y given params θ

$p(\theta|y)$: posterior prob. of params θ given the data observed y

- maximum a-posteriori inference (MAP)
  - simillar to MLE except maximizing $p(\theta|y) \propto p(\theta)p(y|\theta)$
    - same as MLE w/ $p(\theta) \propto 1$
      ↳ * This is an improper prior & is usually not accurate, anyways
  - pros:
    - prior helps regularize
    - can include prior info

- inference summary



- methods of inference: full Bayesian, MAP, MLE

3. Model comparison

- Bayesian decision theory
  - posterior prob → decision

$L(\hat{\theta}, \theta)$: loss function; cost of choosing $\hat{\theta}$ if θ is the real value

$$\hat{\theta}_L = \operatorname{argmin}_{\hat{\theta}} \int \underset{\text{"cost"}}{L(\hat{\theta}, \theta)} \underset{\text{"weight"}}{p(\theta|x,y)} \, d\theta$$

$$= \operatorname{argmin}_{\hat{\theta}} E\left[L(\hat{\theta}, \theta) \mid x, y\right]$$