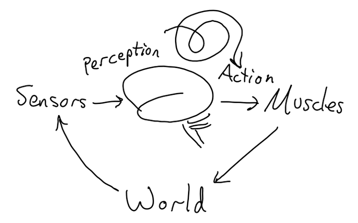
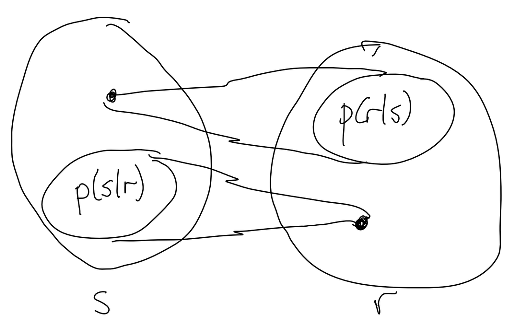
GOALS of Neural Computation



Sensory systems gather information about world



Encoding $p(r|s)$ Decoding $p(s|r)$ Recoding $p(r'|r)$ Acting $p(a|r')$

$s \to r \to r' \to a$
θ memory/learning

Next ½ semester:

Goals of computation (today)

What networks do
- linear
- nonlinear
- stochastic

How networks compute
- feedforward (linear, nonlinear)
- recurrent
- probabilistic

Encoding

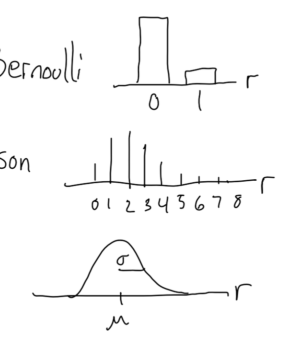
Responses r generated probabilistically

Single neuron examples

$p(r|s) = \begin{cases} r=1 & \lambda \\ r=0 & 1-\lambda \end{cases}$ Bernoulli

$p(r|s) = e^{-\lambda}\lambda^r/r!$ Poisson

$p(r|s) = \dfrac{e^{-\frac{r-\mu}{2\sigma^2}}}{\sqrt{2\pi\sigma^2}}$ Gaussian

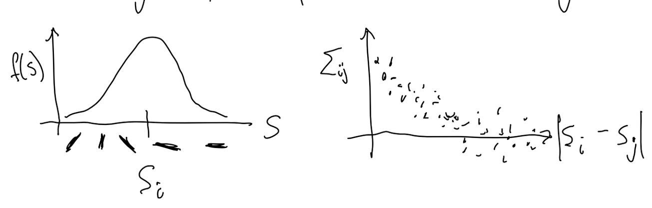


Experiments: choose s, measure/model r

often compute $\langle r|s\rangle = f(s)$ = tuning curves

$\langle \delta r \delta r^T | s\rangle = \Sigma(s)$ = noise correlations

more generally, find $p(r|s)$, but this is high-D



Why is this useful?
- nontrivial tuning suggests s is important for these neurons
- try to reconstruct s?
- compute posterior $p(s|r)$?

Limitations: restricted s
- unnatural context (eg. not full image)
- static
- maybe not the most relevant feature

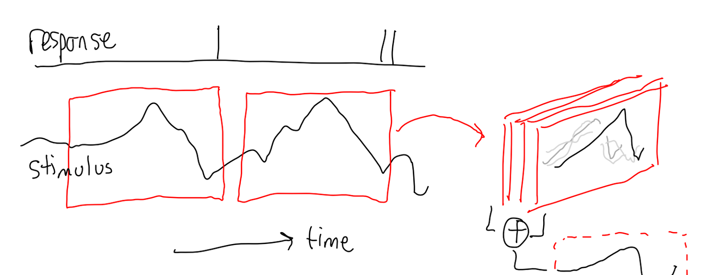
Receptive fields

Generalize tuning to higher-D: $p(r|I)$

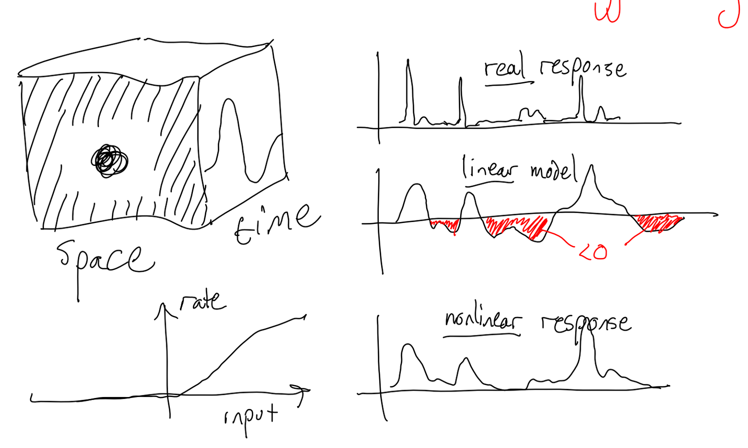
Problem: even for one neuron, this is huge.

Solution: use model for moments

$p(r|I) \to \langle I|r\rangle$ = STA (linear model)
$ \to \langle \delta I \delta I^T|r\rangle$ = STC (quad model)



STA = Spike Triggered Avg



Many more complex encoding models

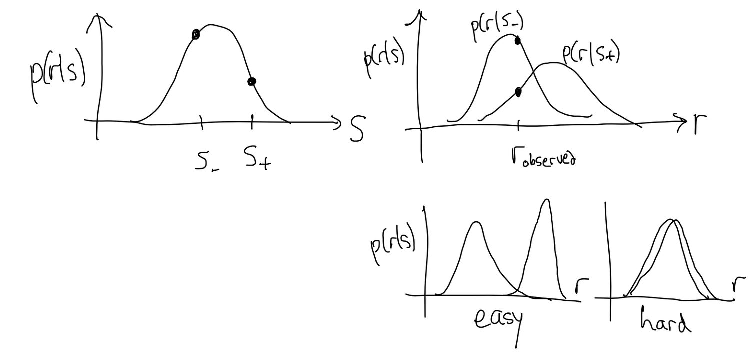
Decoding

What does the decoder know?
- prior distribution of s? Bayesian
- only knows $p(r|s)$? Frequentist ← first

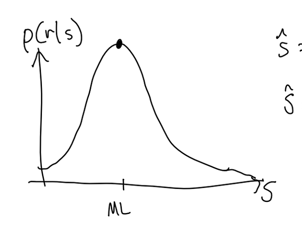
Brain needs to interpret $r \Rightarrow p(r|s)$ likelihood

How to characterize quality of $p(r|s)$?
{discriminate, estimate, calculate uncertainty}

- discriminate: $p(r|s_+) \gtrless p(r|s_-)$



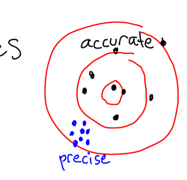
- estimate: $\hat{s} = \arg\max_s p(s|r)$ (Max Likelihood)
  $\hat{s} = w \cdot r + b$ (linear estimator)
  $\hat{s} = g(r)$ (your brain's favorite fcn)

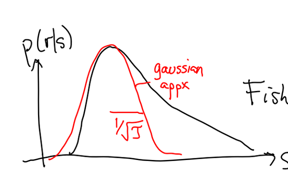


How good is the estimator? $p(\hat{s}|s)$ vs. $p(r|s)$

accuracy: $\langle \hat{s}|s\rangle \neq s \Rightarrow$ bias

precision: $\sigma^2_{\hat{s}} = \langle \delta\hat{s}^2|s\rangle$

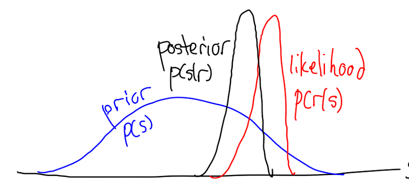




Fisher Info: $J_{\hat{s}} = 1/\sigma^2_{\hat{s}|s}$ (unbiased)
vs. $J_r = 1/\sigma^2_{r|s}$ "

Now: use prior knowledge $p(s)$

posterior $p(s|r) \propto p(r|s)\,p(s)$ (Bayes' rule)



Can still measure variance $\sigma^2_{s|r}$

What is average quality over $p(s)$?

Natural measure is Mutual (Shannon) Information

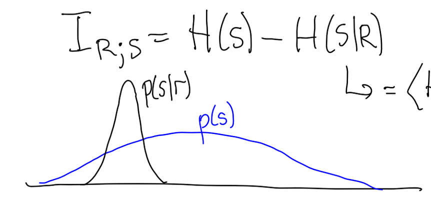
Information for one event with prob. $= p$:

$I(p) \geq 0$
$I(1) = 0$
$* \; I(p_1 \wedge p_2) = I(p_1) + I(p_2)$ ← when independent
$I(p)$ is continuous
$\Rightarrow I(p) = -\log p$

Average information $H(p(s)) = \langle -\log p\rangle_{p(s)}$
"Entropy" $= -\sum_s p(s) \log p(s)$

If measurements are noisy, then info is change in uncertainty:

$I_{R;S} = H(S) - H(S|R)$
$\quad \hookrightarrow = \langle H(p(s|r))\rangle_{p(s|r)}$



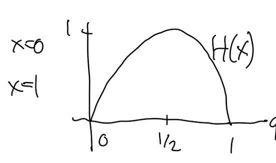
Mutual info, because

$I_{R;S} = H(R) - H(R|S)$
$= I_{S;R} = H(S) - H(S|R)$
$= H(S) + H(R) - H(S,R)$

Examples: $p(x) = \begin{cases} \frac{1}{2} & x=0 \\ \frac{1}{2} & x=1 \end{cases} \Rightarrow H(X) = 1$

$p(x) = \frac{1}{N} \; \forall x \in \{1,\dots,N\} \Rightarrow H(X) = ?$

$p(x) = \begin{cases} q & x=0 \\ 1-q & x=1 \end{cases}$



p(x,y)	0	1
0	1/2	
1		1/2

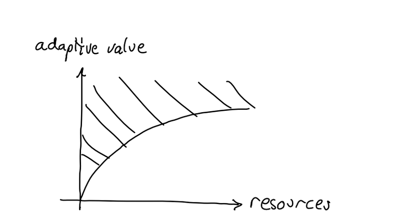
p(x,y)	0	1
0	1/4	1/4
1	1/4	1/4

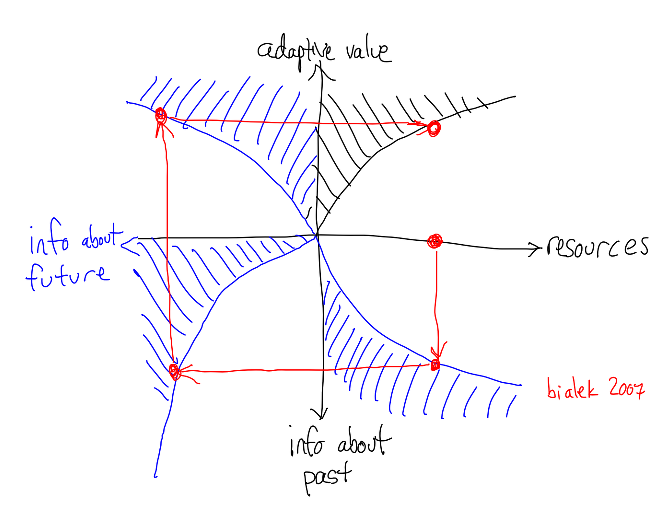
p(x,y)	0	1
0		1/2
1	1/2	

p(x,y)	0	1
0	3/8	1/8
1	1/8	3/8

$H(X,Y) = -2\left(\frac{3}{8}\log\frac{3}{8} + \frac{1}{8}\log\frac{1}{8}\right) \approx 1.8$

Information about what?





Encoded information should be useful, ie predictive: $I_{past;future}$

This can be about slowly-changing things, and can be represented in either activity or synapses...

Throw away the rest of the information
This is COMPUTATION.

Brain does not operate on $\hat{s}$ or $p(\hat{s})$ or $p(s|r)$. It operates on r

$r \to r'$ "Recoding"

That is our main topic.

Summary

Encoding ↑ Descriptive

Decoding ↑ Needed for experiments

Recoding ↑ What the brain does

$p(s|\hat{s}(r))$ vs $p(s|\hat{s}'(r'))$

Information always decreases $I_{s;r} \geq I_{s;r'}$

But accessible Info increases! $I_{s;\hat{s}(r)} < I_{s;\hat{s}'(r')}$

This is because estimators have limited capability and information is represented in a complicated way.

This is neural computation!