Dall-E

EM Algorithm

Applied Machine Learning
Derek Hoiem

Today’s Class: EM Algorithm

I will describe three problems. Think about what they have in common.

“Bad Annotators” Problem

You want to train an algorithm to predict whether a photograph is attractive. You collect annotations from Mechanical Turk. Some annotators try to give accurate ratings, but others answer randomly.

Challenge: Determine which people to trust and the average rating by accurate annotators.

Annotator Ratings

10
8
9
2
8

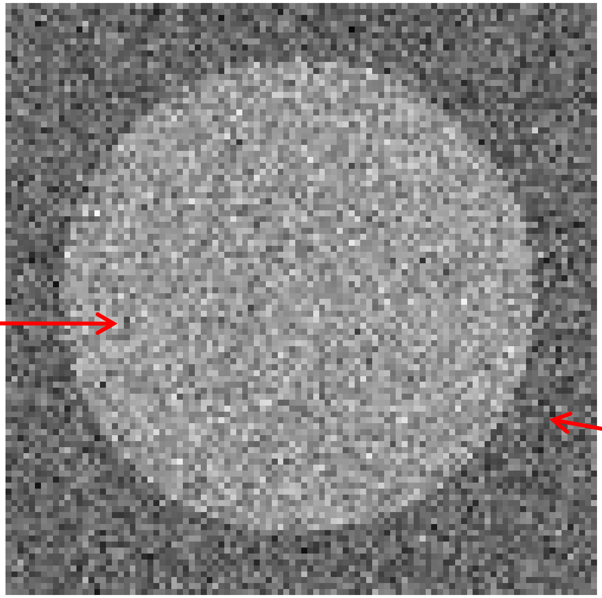
Photo: Jam343 (Flickr)

Foreground/Background Segmentation

You are given an image and want to assign foreground/background pixels.

Challenge: Segment the image into figure and ground without knowing what the foreground looks like in advance.

Foreground



Background

Topic Models

Documents have a “topic” that is predictive of the words

Challenge: We don’t know what the topics are, or what distribution of words each has

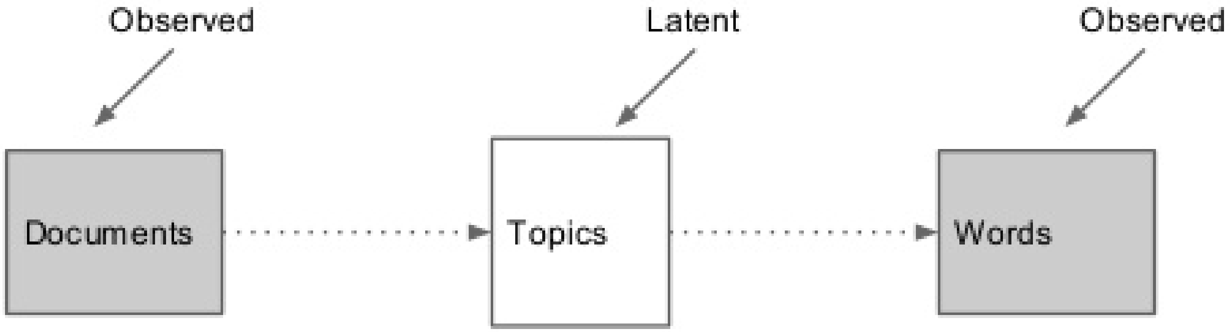


Fig source

What do these problems have in common?

1. We think there is some underlying factor that is not known
   – Whether annotator is good or bad
   – Whether pixel is in foreground or background
   – Topic of a document

2. We have some model for the probability of data given that underlying factor

3. But we don’t know the parameters of the model

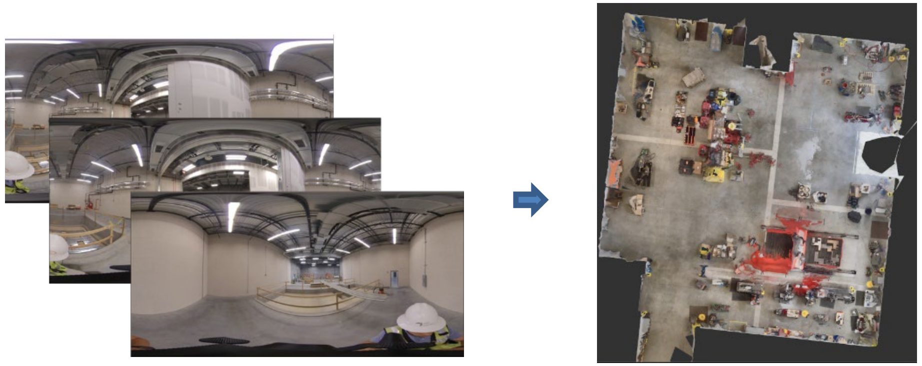
These are “missing data” or “latent variable” problems – a critical piece of information is not observed

Today's Class

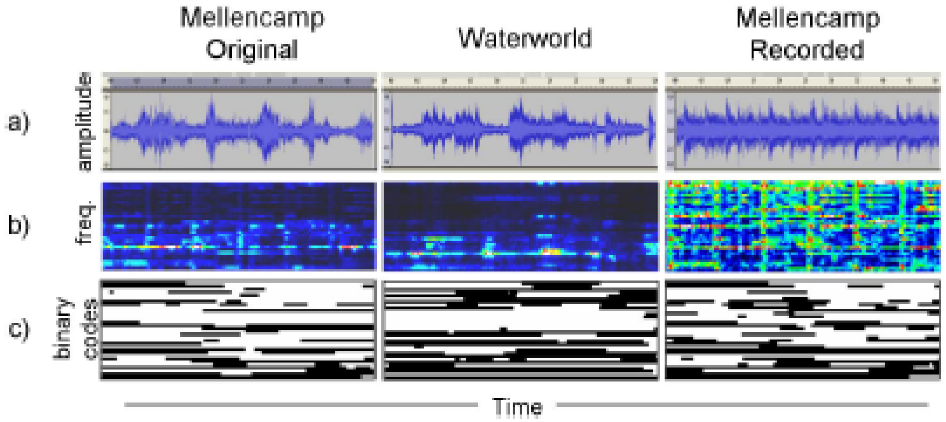
- Examples of problems with hidden or latent variables
  - Untrustworthy annotators
  - Pixel segmentation
  - Topic models
- Background
  - Maximum Likelihood Estimation
  - Probabilistic Inference
- Dealing with Latent Variables (latent = hidden, not observed)
  - EM algorithm, Bad annotator problem
  - Hard EM

I have used EM in research and practice many times, e.g.

- Given multiple images and scene geometry, estimate true color of each floor map pixel
  - Latent variables: which pixels are occluded



- For an audio clip of music with background noise, which extracted sound signatures are due to music vs background noise?
  - Latent variables: whether each extracted sound signature at a given time is due to background or music



- Mixture of Gaussian probability model (next class)

Bad Annotator Problem

You hire annotators to label attractiveness of images. Some annotators do their best. Some are "bad" and assign random scores.

Goal: We want to estimate the average score of good annotators for each image

Assumptions

1. Bad annotators are always bad. Good annotators are always good.
2. For each image i, the scores from good annotators follow a Gaussian distribution $s_{ia} \sim N(\mu_i, \sigma^2)$

$$P(s_{ia}|z_a = 1) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2}\frac{(s_{ia} - \mu_i)^2}{\sigma^2}\right)$$

3. The scores from bad annotators always follow a uniform distribution

$$P(s_{ia}|z_a = 0) = 1 \quad \text{(scores range from 0 to 1)}$$

Notation

$s_{ia} \in [0,1]$: score for image i by annotator *a*

$z_a \in \{0,1\}$: whether annotator a is good

μ_i: true mean score for image i

σ: standard deviation of true scores, same for each image

π_z: $P(\mathrm{z_a} = 1)$, prior probability that annotator is good

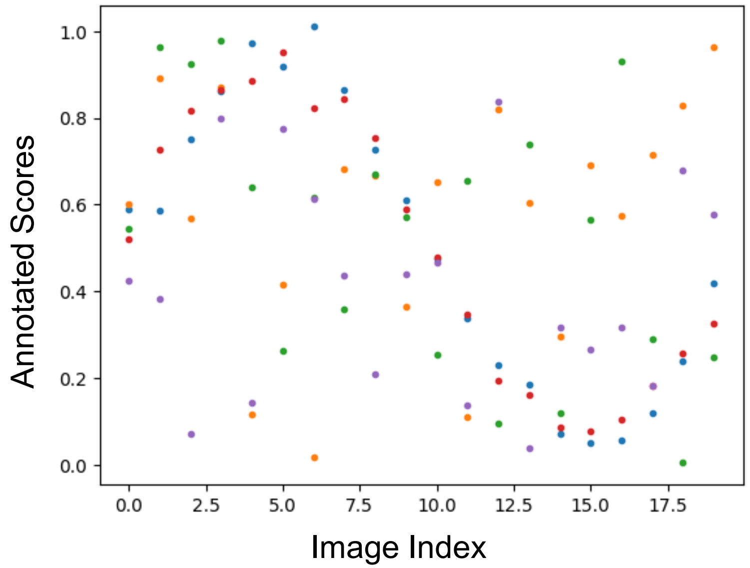
Latent Variable Problems: Bad Annotator

Challenge: Figure out which annotators are good and estimate the true mean score for each image

Three steps:

1. If we knew which annotators were good, how would we estimate the score distribution for each image?
2. Given the distribution parameters, how do we compute the likelihood that an annotator is good?
3. How can we get annotator labels and score models at once?

i=0	i=1	i=2	…
0.59	0.88	0.7	
0.52	0.96	0.71	
0.50	0.88	0.52	
0.39	0.94	0.65	
0.65	0.69	0.46	



Maximum Likelihood Estimation

1. If we knew which annotators were good, how would we estimate the score distribution for each image?

MLE:

data

$$\underline{s}_i = \{s_{i0} \dots s_{im}\}$$

parameters

Solve for parameters that maximize the data likelihood

$$\hat{\theta} = \underset{\theta}{\operatorname{argmax}}\ p(\underline{s}_i \mid \theta)$$

Scores are independent of each other, given the model

$$\hat{\theta} = \underset{\theta}{\operatorname{argmax}} \prod_a p(s_{ia} \mid \theta)$$

Model is Gaussian

$$p(s_{ia} \mid \mu_i, \sigma) = \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left(-\frac{1}{2}\frac{(s_{ia}-\mu_i)^2}{\sigma^2}\right)$$

Solving for mean

$$p(s_{ia} | \mu_i, \sigma) = \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left(-\frac{1}{2}\frac{(s_{ia}-\mu_i)^2}{\sigma^2}\right)$$

Easier to take derivative of sum of logs than a product, and f(x) and log(f(x)) are always maximized by the same x

$$\underset{\theta}{\operatorname{argmax}} \prod_a p(s_{ia} | \theta)$$

To find max wrt a variable, set the partial derivative wrt that variable to 0

$$\frac{\partial}{\partial \mu} \sum_a \left[\log\frac{1}{\sqrt{2\pi}} + \log\frac{1}{\sigma} - \frac{1}{2}\frac{(s_{ia}-\mu_i)^2}{\sigma^2}\right] = 0$$

Do the math

$$\sum_a (s_{ia}-\mu_i)/\sigma^2 = 0 \Rightarrow \sum_a s_{ia} - \sum_a \mu_i = 0$$

Solve. M is number of “a”s

$$\mu_i = \sum_a s_{ia} / M$$

Solving for standard deviation

Now take partial derivative wrt sigma. Since sigma is the same for all images, we are now summing over all images and annotations

$$\frac{\partial}{\partial \sigma} \sum_i \sum_a \left[\log \frac{1}{\sqrt{2\pi}} + \log \frac{1}{\sigma} - \frac{1}{2} \frac{(s_{ia} - \mu_i)^2}{\sigma^2} \right] = 0$$

$$\sum_i \sum_a \left[-\frac{1}{\sigma} + \frac{(s_{ia} - \mu_i)^2}{\sigma^3} \right] = 0 \Rightarrow \sum_{i,a} -\sigma^2 + \sum_{i,a} (s_{ia} - \mu_i)^2 = 0$$

Solution is average squared difference from mean

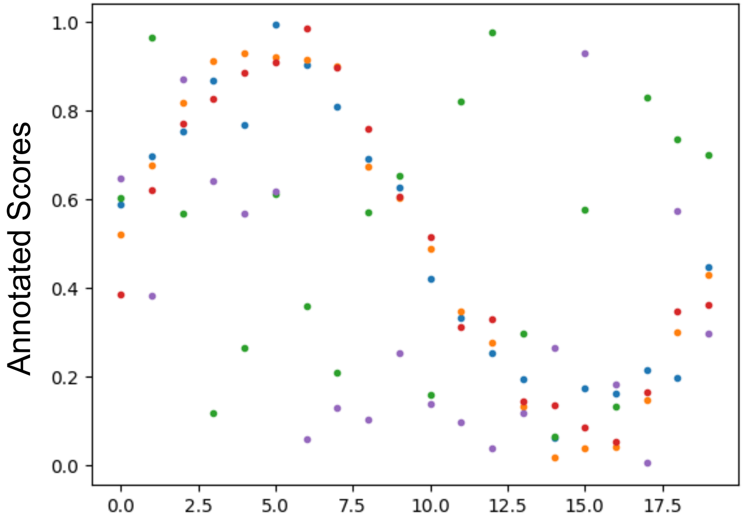
$$\sigma^2 = \sum_{i,a} (s_{ia} - \mu_i)^2 / (M \cdot N)$$

Applying MLE in code

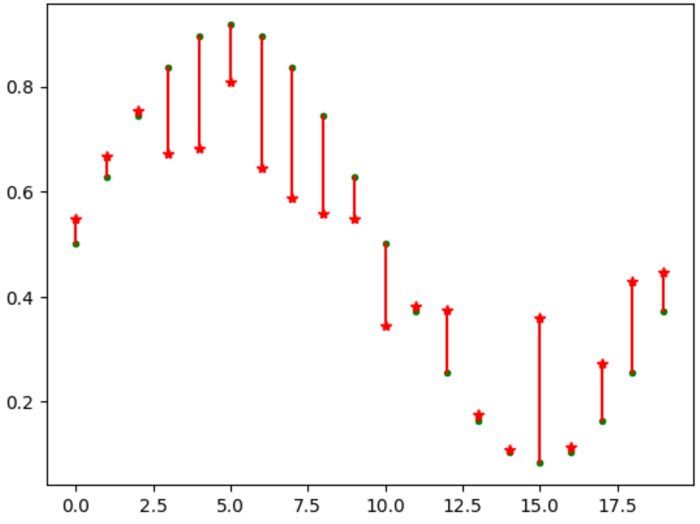
$$\mu_i = \sum_a s_{ia} / M$$

$$\sigma^2 = \sum_{i,a} (s_{ia} - \mu_i)^2 / (M \cdot N)$$

```
# initialize by assuming that all scores are good
score_mean = scores.mean(axis=1).reshape((len(scores), 1))  # mu_i
score_std = np.sqrt(np.sum((scores-score_mean)**2, axis=None)/N/M)  # sigma
```

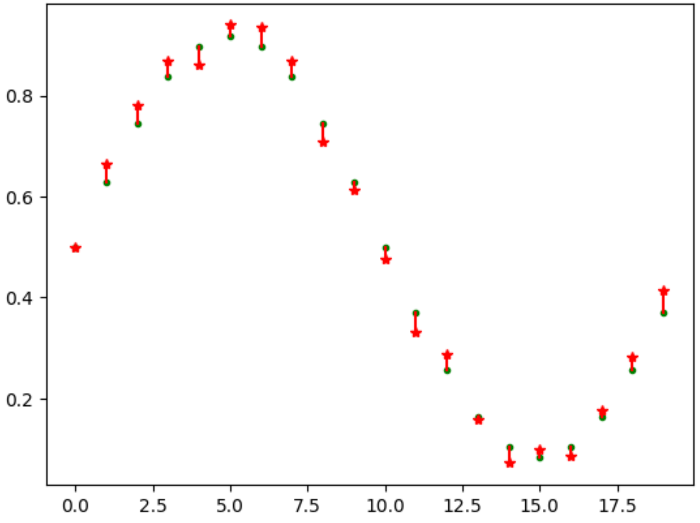


Assuming all the scores are good (when they are not)



(Green = true; red = prediction)

If we knew somehow which annotators are good

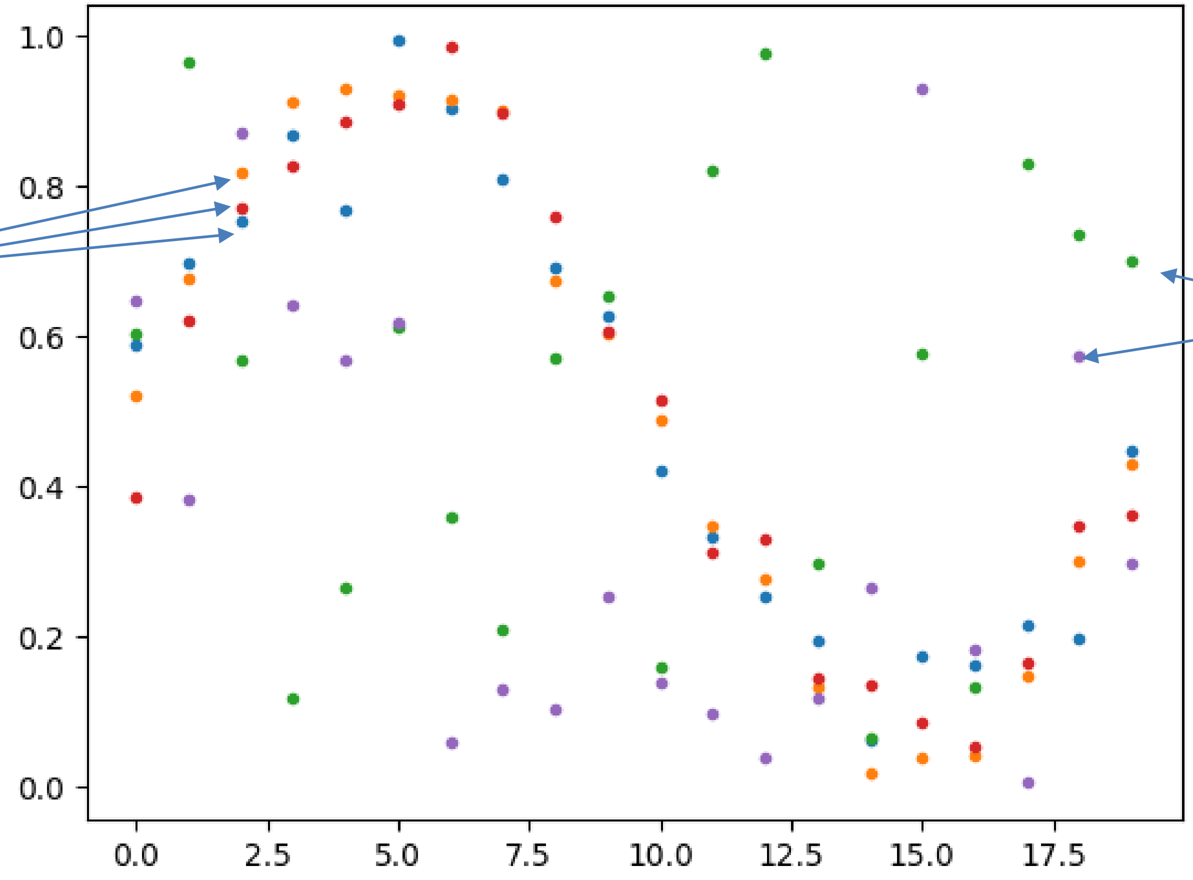


(Green = true; red = prediction)

Probabilistic Inference

2. Given the distribution parameters, how do we compute the likelihood that an annotator is good?

Good annotators
(red, orange, blue)



Bad annotators
(green, purple)

Note: In this example, I know what is good and bad because I made up the scores, but otherwise it wouldn't be obvious

Probabilistic Inference (general case)

Given the model parameters, compute the likelihood that a particular model generated a sample

component or label

$$p(z_n = m \mid x_n, \theta)$$

z_n is the unknown label of data point x_n

General strategy: We know $p(x_n|z_n = 0, \theta)$ and $p(x_n|z_n = 1, \theta)$ and $p(z_n = 1|\theta)$

We want to know $p(z_n = m \mid x_n, \theta)$

Use probability rules to get from what we know to what we want to know.

Probabilistic Inference (general case)

Given the model parameters, compute the likelihood that a particular model generated a sample

component or label

$$p(z_n = m \mid x_n, \theta) = \frac{p(z_n = m, x_n \mid \theta_m)}{p(x_n \mid \theta)}$$

Rule of conditional probability

Probabilistic Inference (general case)

Given the model parameters, compute the likelihood that a particular model generated a sample

component or label

$$p(z_n = m \mid x_n, \theta) = \frac{p(z_n = m, x_n \mid \theta_m)}{p(x_n \mid \theta)}$$

$$= \frac{p(z_n = m, x_n \mid \theta_m)}{\sum_k p(z_n = k, x_n \mid \theta_k)}$$

Law of total probability

Probabilistic Inference (general case)

Given the model parameters, compute the likelihood that a particular model generated a sample

component or label

$$p(z_n = m \mid x_n, \theta) = \frac{p\left(z_n = m, x_n \mid \theta_m\right)}{p\left(x_n \mid \theta\right)}$$

$$= \frac{p\left(z_n = m, x_n \mid \theta_m\right)}{\sum_k p\left(z_n = k, x_n \mid \theta_k\right)}$$

$$= \frac{p\left(x_n \mid z_n = m, \theta_m\right) p\left(z_n = m \mid \theta_m\right)}{\sum_k p\left(x_n \mid z_n = k, \theta_k\right) p\left(z_n = k \mid \theta_k\right)}$$

Chain rule of probability

Example: Inference for Annotator Labels

$$p(z_a = 1|s_a, \theta) = \frac{p(s_a|z_a = 1, \theta)p(z_a = 1, \theta)}{p(s_a|z_a = 1, \theta)p(z_a = 1, \theta) + p(s_a|z_a = 0, \theta)p(z_a = 0, \theta)}$$

$p(s_a|z_a = 1, \theta) = \prod_i N(s_{ia}, \mu_i, \sigma)$ (normal pdf)

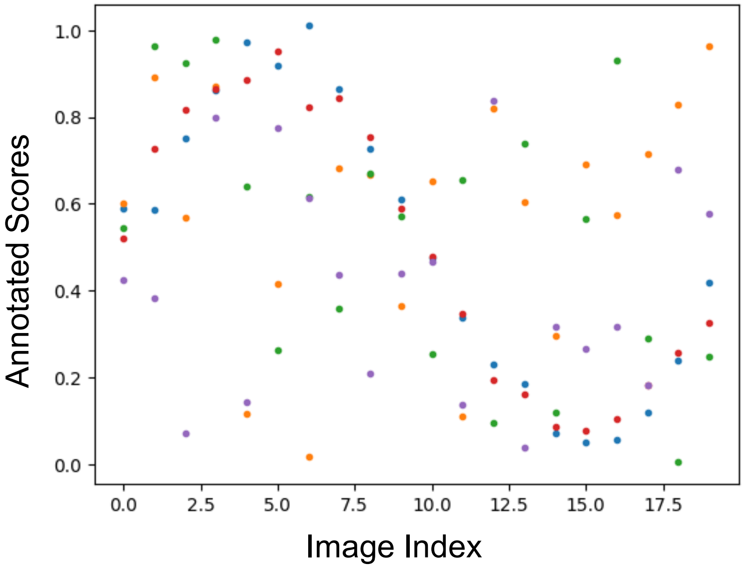
$p(s_a|z_a = 0, \theta) = \prod_i 1$ (uniform)

$p(z_a = 1, \theta) = \pi_z$ (prior)

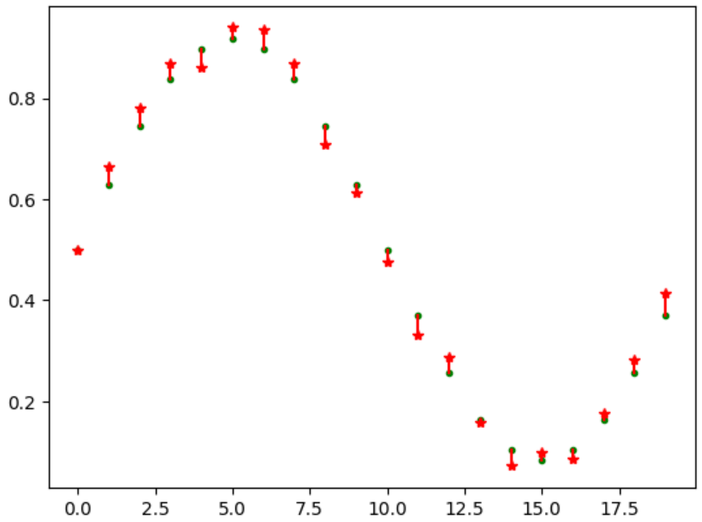
```
p_good = np.zeros((5,1))  # w_a = P(z_a=1 | scores, theta_t)
for a in range(M):
  p_s_good = pz  # P(s_ia | z=1, mu_i, std)P(z_a=1)
  p_s_bad = 1-pz # P(s_ia | z=0)P(z_a=0)
  for i in range(N):
    p_s_good *= 1/np.sqrt(2*np.pi)/score_std * np.exp(-1/2 * (scores[i,a]-score_mean[i])**2/score_std**2)
    p_s_bad *= 1 # uniform in range [0, 1]
  p_good[a] = p_s_good / (p_s_good + p_s_bad)
```

Dealing with Latent Variables

3. How can we get annotator labels and score models at once?



Estimated scores



(Green = true; red = prediction)

Good annotators: 0, 1, 3

2 minute break

- After the break:
  - Intuitive solution to solving for parameters and latent variables
  - Derive that the intuitive solution is correct
  - Demo for the untrustworthy annotator problem

Dall-E3 The three bears, little red riding hood, her grandmother, and a dragon at a tea party with a huge cake in a hotel with people in the background. Digital art.

A cat looks on in shock as a person on hands and knees eats the cat's food. Digital art.

Simple solution

1. Initialize parameters

2. Compute the probability of each hidden variable given the current parameters

3. Compute new parameters for each model, weighted by likelihood of hidden variables

4. Repeat 2-3 until convergence

Annotator Problem: Simple Solution

1. Initialize parameters
   - Estimate parameters assuming all annotators are good

2. Compute likelihood of hidden variables for current parameters $w_a = p(z_a = 1 | \boldsymbol{s}_a, \boldsymbol{\mu}, \sigma, \pi_z)$

3. Estimate new parameters for each model, weighted by likelihood

$\hat{\mu}_i = \sum_a^1 w_a s_{ia} / \sum_a^1 w_a$ $\hat{\sigma}^2 = \sum_{i,a}^1 (s_{ia} - \mu_i)^2 w_a / \sum_{i,a}^1 w_a$ $\pi_z = \sum_a^1 w_a / M$

Expectation Maximization (EM) Algorithm

Goal: $\hat{\theta} = \underset{\theta}{\operatorname{argmax}} \log\left(\sum_{\mathbf{z}} p(\mathbf{x}, \mathbf{z} \mid \theta)\right)$

Log of sums is intractable

Jensen's Inequality

$$f(\mathrm{E}[X]) \geq \mathrm{E}[f(X)]$$

for concave functions f(x)

So we maximize the lower bound by maximizing the log of sums instead of sum of logs!

Expectation Maximization (EM) Algorithm

Goal: $\hat{\theta} = \underset{\theta}{\operatorname{argmax}} \log\left(\sum_{\mathbf{z}} p(\mathbf{x}, \mathbf{z} \mid \theta)\right)$

1. E-step: compute

$$\mathrm{E}_{z|x,\theta^{(t)}}\left[\log(p(\mathbf{x}, \mathbf{z} \mid \theta))\right] = \sum_{\mathbf{z}} \log(p(\mathbf{x}, \mathbf{z} \mid \theta)) p\left(\mathbf{z} \mid \mathbf{x}, \theta^{(t)}\right)$$

2. M-step: solve

$$\theta^{(t+1)} = \underset{\theta}{\operatorname{argmax}} \sum_{\mathbf{z}} \log(p(\mathbf{x}, \mathbf{z} \mid \theta)) p\left(\mathbf{z} \mid \mathbf{x}, \theta^{(t)}\right)$$

Expectation Maximization (EM) Algorithm

log of expectation of p(x|z) over p(z)

Goal: $\hat{\theta} = \underset{\theta}{\operatorname{argmax}} \log\left(\sum_{\mathbf{z}} p(\mathbf{x}, \mathbf{z} \mid \theta) \right)$ $\qquad f\left(\mathrm{E}[X]\right) \geq \mathrm{E}\left[f(X)\right]$

1. E-step: compute

expectation of log of P(x|z) over estimated P(z)

$$\mathrm{E}_{z|x,\theta^{(t)}}\left[\log\left(p(\mathbf{x}, \mathbf{z} \mid \theta)\right)\right] = \sum_{\mathbf{z}} \log\left(p(\mathbf{x}, \mathbf{z} \mid \theta)\right) p\left(\mathbf{z} \mid \mathbf{x}, \theta^{(t)}\right)$$

2. M-step: solve

$$\theta^{(t+1)} = \underset{\theta}{\operatorname{argmax}} \sum_{\mathbf{z}} \log\left(p(\mathbf{x}, \mathbf{z} \mid \theta)\right) p\left(\mathbf{z} \mid \mathbf{x}, \theta^{(t)}\right)$$

EM for Annotator Problem: **E-Step**

μ_i, σ^2, π_z

E-step: $E_{z|s,\theta^{(t)}}\left(\log P(\underline{s}, \underline{z} \mid \theta)\right) = \sum_{\underline{z}} \log P(\underline{s}, \underline{z} \mid \theta) P(\underline{z} \mid \underline{s}, \theta^{(t)})$

Use $\hat{\theta}$ as current estimate of θ

$\sum_{\underline{z}} \log P(\underline{s}, \underline{z} \mid \theta) P(\underline{z} \mid \underline{s}, \hat{\theta})$

These sums over all combinations of **z** and joint probabilities are complicated!

First, let's expand the joint probabilities

EM for Annotator Problem: E-Step

Write joint as product of individual

Log of product is sum of logs

μ_i, σ^2, π_z

E-step: $E_{Z|S,\theta^{(t)}}\left(\log P(S,Z|\theta)\right) = \sum_{Z} \log P(S,Z|\theta) P(Z|S,\theta^{(t)})$

Use $\hat{\theta}$ as current estimate of θ

$$\sum_{Z} \log P(S,Z|\theta) P(Z|S,\hat{\theta})$$

$$= \sum_{Z} \left[\log \prod_{i,a} P(s_{ia}|z_a,\theta) P(z_a|\theta)\right] P(Z|S,\hat{\theta})$$

$$= \sum_{Z} \left[\sum_{i,a} \log\left(P(s_{ia}|z_a,\theta) P(z_a|\theta)\right)\right] P(Z|S,\hat{\theta})$$

Now I want to deal with this joint **z** and complex sum over **z**

EM for Annotator Problem: E-Step

μ_i, σ^2, π_z

E-step: $E_{z|s,\theta^{(t)}}(\log P(\underline{s},\underline{z}|\theta)) = \sum_{\underline{z}} \log P(\underline{s},\underline{z}|\theta) P(\underline{z}|\underline{s},\theta^{(t)})$

Use $\hat{\theta}$ as current estimate of θ

$$\sum_{\underline{z}} \log P(\underline{s},\underline{z}|\theta) P(\underline{z}|\underline{s},\hat{\theta})$$

$$= \sum_{\underline{z}} \left[\log \prod_{i,a} P(s_{ia}|z_a,\theta) P(z_a|\theta)\right] P(\underline{z}|\underline{s},\hat{\theta})$$

$$= \sum_{\underline{z}} \left[\sum_{i,a} \log(P(s_{ia}|z_a,\theta) P(z_a|\theta))\right] P(\underline{z}|\underline{s},\hat{\theta})$$

Rearrange sums

$$= \sum_{i,a} \sum_{\underline{z}} [\qquad]$$

Make it clear that sum over z is a series of sums

$$= \sum_{i,a} \sum_{z_0} \cdots \sum_{z_m} [\qquad]$$

Rearrange again, pulling out the sum over z_a

only depends on z_a

$$= \sum_{i,a} \sum_{z_a} [\log P(s_{ia}|z_a,\theta) + \log P(z_a|\theta)] \underbrace{\sum_{z_{\not a}} P(\underline{z}|\underline{s},\hat{\theta})}_{P(z_a|\underline{s},\hat{\theta})}$$

This lets me marginalize out the other z's!

EM for Annotator Problem: E-Step

$$\text{E-step: } E_{Z|S,\theta^{(t)}}\left(\log P(\underline{S},\underline{Z}|\theta)\right) = \sum_{\underline{Z}} \log P(\underline{S},\underline{Z}|\theta) P(\underline{Z}|S,\theta^{(t)})$$

μ_i, σ^2, π_z

Use $\hat{\theta}$ as current estimate of θ

$$\sum_{\underline{Z}} \log P(\underline{S},\underline{Z}|\theta) P(\underline{Z}|\underline{S},\hat{\theta})$$

$$= \sum_{\underline{Z}} \left[\log \prod_{i,a} P(S_{ia}|Z_a,\theta) P(Z_a|\theta)\right] P(\underline{Z}|\underline{S},\hat{\theta})$$

$$= \sum_{\underline{Z}} \left[\sum_{i,a} \log\left(P(S_{ia}|Z_a,\theta) P(Z_a|\theta)\right)\right] P(\underline{Z}|\underline{S},\hat{\theta})$$

$$= \sum_{i,a} \sum_{\underline{Z}} \left[\qquad \right]$$

$$= \sum_{i,a} \sum_{Z_0} \cdots \sum_{Z_m} \left[\qquad \right]$$

only depends on Z_a

$$= \sum_{i,a} \sum_{Z_a} \left[\log P(S_{ia}|Z_a,\theta) + \log P(Z_a|\theta)\right] \underbrace{\sum_{\underline{Z}_{\not{a}}} P(\underline{Z}|\underline{S},\hat{\theta})}_{P(Z_a|\underline{S},\hat{\theta})}$$

$Z=1$

$$= \sum_{i,a} \left[\log \frac{1}{\sqrt{2\pi}\sigma} - \frac{1}{2}\frac{(S_{ia}-\mu_i)^2}{\sigma^2} + \log \pi_z\right] P(Z_a=1|\underline{S},\hat{\theta})$$
$$+ \left[\underbrace{\log(1)}_{0} + \log(1-\pi_z)\right] P(Z_a=0|\underline{S},\hat{\theta})$$

Now, plug in the functions for score probabilities of good and bad annotators

$$P(Z_a=1|\underline{S},\hat{\theta}) = P(Z_a=1|\underline{S}_a,\hat{\theta})$$
$$= P(Z_a=1,\underline{S}_a|\hat{\theta}) / P(\underline{S}_a|\hat{\theta})$$
$$= P(\underline{S}_a|Z_a=1,\hat{\theta}) P(Z_a=1|\hat{\theta}) \Big/ \left[P(\underline{S}_a|Z_a=1,\hat{\theta}) P(Z_a=1|\hat{\theta}) + P(\underline{S}_a|Z_a=0,\hat{\theta}) P(Z_a=0|\hat{\theta})\right]$$

And calculate how to get P(z|s), as we did before

EM for Annotator Problem: E-Step

Writing out the inference computations

$$P(z_a=1\mid \underline{s},\hat{\theta}) = P(z_a=1\mid \underline{s}_a,\hat{\theta})$$

$$= P(z_a=1,\underline{s}_a\mid\hat{\theta}) / P(\underline{s}_a\mid\hat{\theta})$$

$$= P(\underline{s}_a\mid z_a=1,\hat{\theta})P(z_a=1\mid\hat{\theta}) \Big/ \left[P(\underline{s}_a\mid z_a=1,\hat{\theta})P(z_a=1\mid\hat{\theta}) + P(\underline{s}_a\mid z_a=0,\hat{\theta})P(z_a=0\mid\hat{\theta})\right]$$

$$P(\underline{s}_a\mid z_a=1,\hat{\theta}) = \prod_i P(s_{ia}\mid z_a=1,\hat{\theta}) = \prod_i N(s_{ia};\hat{\mu}_i,\hat{\sigma}_i^2)$$

$$P(z_a=1\mid\hat{\theta}) = \pi_z \qquad P(z_a=0\mid\hat{\theta}) = 1-\pi_z$$

$$P(\underline{s}_a\mid z_a=0,\hat{\theta}) = \prod_i P(s_{ia}\mid z_a=0,\hat{\theta}) = 1$$

Let $P(z_a=1\mid\underline{s},\hat{\theta}) = w_a$

EM for Annotator Problem: **M-Step**

Let $P(z_a=1 \mid \underline{s}, \hat{\theta}) = w_a$

Calculate parameters that maximize the expression from the E-step, given our current estimates of P(z|s)

This is very similar to the MLE derivation

M-Step: find parameters that maximize expected log likelihood

$$\frac{\partial}{\partial \mu_i} \sum_{i,a} \left[\log \frac{1}{\sqrt{2\pi}\sigma} - \frac{1}{2}\frac{(s_{ia}-\mu_i)^2}{\sigma^2} + \log \pi_z \right] w_a + \log(1-\pi_z)(1-w_a) = 0$$

$$\sum_a \left[\frac{(s_{ia}-\mu_i)}{\sigma^2}\right] w_a = 0 \Rightarrow \hat{\mu}_i = \sum_a w_a s_{ia} / \sum_a w_a$$

$\hat{\mu}_i$

$$\frac{\partial}{\partial \sigma} \sum_{i,a} \left[\log \frac{1}{\sigma} - \frac{1}{2}\frac{(s_{ia}-\mu_i)^2}{\sigma^2}\right] w_a = 0$$

$$\sum_{i,a} \left[-\frac{1}{\sigma} + \frac{(s_{ia}-\mu_i)^2}{\sigma^3}\right] w_a = 0 \Rightarrow \sum_{i,a}\left[-\sigma^2 + (s_{ia}-\mu_i)^2\right] w_a = 0$$

$$\sigma^2 = \sum_{i,a} (s_{ia}-\mu_i)^2 w_a / \sum_{i,a} w_a$$

$\hat{\sigma}$

$$\frac{\partial}{\partial \pi_z} \sum_{i,a} \log \pi_z w_a + \log(1-\pi_z)(1-w_a) = 0$$

$$\sum_{i,a} \frac{1}{\pi_z} w_a - \frac{1}{1-\pi_z}(1-w_a) = 0 \Rightarrow \sum_{i,a} (1-\pi_z) w_a - \pi_z (1-w_a) = 0$$

$$\sum_{i,a} \left[w_a - \pi_z w_a - \pi_z + \pi_z w_a\right] = 0$$

$$\pi_z = \sum_{i,a} w_a / \sum_{i,a} 1 = \sum_a w_a / M$$

$\widehat{\pi_z}$

EM Annotator Problem Demo

https://colab.research.google.com/drive/1sutnFg-xe-IjgiY8qAJt5USf2MEBZS2L?usp=sharing

EM Algorithm

- Maximizes a lower bound on the data likelihood at each iteration

- Each step increases the data likelihood
  - Converges to *local maximum*

- Common tricks to derivation
  - Find terms that sum or integrate to 1
  - Lagrange multiplier to deal with constraints

- Although the derivation is long, it pretty much always boils down to iteratively
  1. Estimating likelihood of latent variables given parameters
  2. Computing estimates of parameters that are weighted by the latent variable likelihoods

“Hard EM”

- Same as EM except compute z* as most likely values for hidden variables

- K-means is an example

- Advantages
  – Simpler: can be applied when cannot derive EM
  – Sometimes works better if you want to make hard predictions at the end
- But
  – Generally, pdf parameters are not as accurate as EM

Notes about homeworks

- HW 2: due Feb 19
  - Problem 1: PCA (lecture Jan 30)
  - Problem 2: Linear Classification, parameter selection (lectures Feb 1, Feb 6)
  - Problem 3: Linear Regression (lecture Feb 1)
  - Stretch goals: optional, up to 60 points

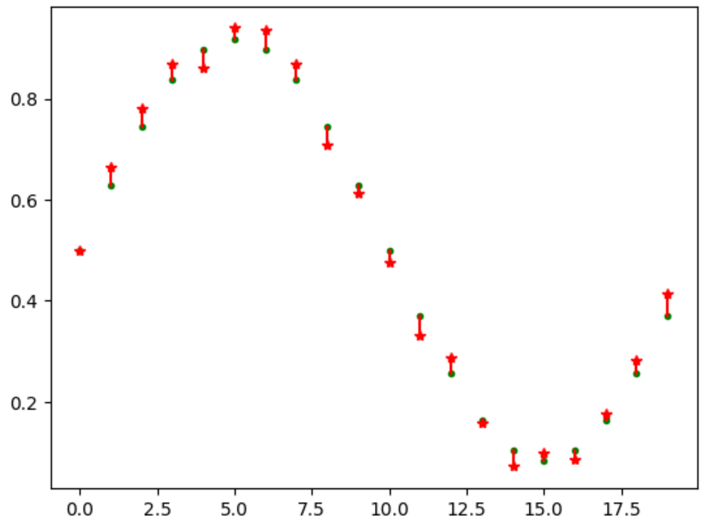
- HW 3: due Mar 4
  - Problem 1: Probability estimation and inference (lectures Feb 8-15)
  - Problem 2: Robust Estimation of Salary Data (lectures Feb 13, Feb 15, Feb 20)
    - Some salaries are “true reports” coming from some distribution based on university and experience; others are “fake reports” that are generated uniformly at random
  - Stretch goals: optional, up to 60 points

- Remember, you only need a total of 500 (3 credit) or 625 (4 credit) points, which can be earned through HWs, final project, and participation

What to remember

- EM is a widely applicable algorithm to solve for latent variables and parameters that make the observed data likely
- While derivation is long and somewhat complicated, the application is simple
- We'll see other examples of EM use in mixture of Gaussian and topic models

Estimated scores



(Green = true; red = prediction)

Good annotators: 0, 1, 3

Next class

- Estimating probability density functions