**Message passing and Factor graphs** (BIG IDEA: messages represents marginalized probability (to different degrees)

|  |  |  |
| --- | --- | --- |
|  |  |  |
|  |
|  |
|  |
|  |  |  |
|  |

**K-means,** **autoencoders** error function

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | | |  | |
| PCA  max var |  | | |  |
| PCA  min err | given minimize | After minimization, we obtain    Find m eigenvectors with max eig values | | Decompose S =XXT X is N by D (N<D)    Reduce from O(D3) to O(N3) |

Prediction (distribution) with entire weight space (not a single MLE/MAP)

|  |  |
| --- | --- |
|  |  |

Conjugate prior method for finding

|  |  |
| --- | --- |
| If α→0, |  |

Model Comparison

|  |  |  |  |
| --- | --- | --- | --- |
| Same prori thus derive **bayes factor** to assess models | Model evidence/Marg. Liklihood of P(w|D,M)  Predictive mixture distribution | Probability of data approx using MAP | 2 compete terms first term is height second term is width; optimum is max (height + width) on average |
|  |  |  |  |

Conjugate Priors bionomial beta, Assume bionomial data liklihood and given model, model evidence

|  |  |
| --- | --- |
|  |  |

**1st EM intution ; MoG-----**find to maximize **;** parmeters in second column is also the result of M-step on ECDL (Expected Complete Data Log-liklihood)

|  |  |  |
| --- | --- | --- |
| ; | ; | Notice depends on, which depends on →we need pick an initial recursion until convergence! |

**2nd intuition of EM;** still want to max log , this time **simply** maximize , but we don’t have z so opimitze ECDL(which is lower bound)

|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  | |
| thus | | | M-step for CDL:  **;** ;  ; |

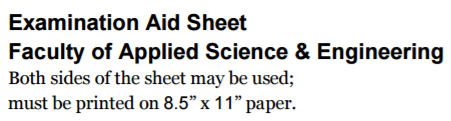
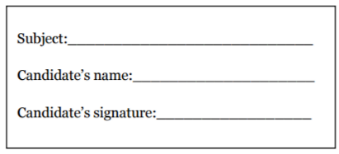
**EM MORAL: write CDL p(x,z)=logp(x|z,θ)p(z|θ ), differenciate wrt θ and set derivative to 0, replace indicator functions with responsibility**

**Naïve Bayes Classifier**; Instead of generate using a mixture of in MoG, we classify input to latent cause by assuming C.I. (say z takes on {0,1})

|  |  |
| --- | --- |
| Assume:  ; |  |

**Markov Assumption**---future only depends on present

|  |  |
| --- | --- |
| Given |  |



**Bayes Net, MRV, Factor graphs**

|  |  |  |
| --- | --- | --- |
|  |  |  |

Bayes ball

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | FALSE |  | FALSE |  | FALSE |  | FALSE |
| TRUE | TRUE | TRUE | TRUE |

**HMM** with α-β recursion O(K^2) ; Objective maximize P(x) by simply maximizing P(X,Z) wrt θ , use EM

|  |  |  |  |
| --- | --- | --- | --- |
| Given |  | |  |
| Sum above over all , we have At  K×K transition matrix  **EM= take deriv. of CDL replace z with γ(zn), replace transition with** | | |  |
| CDL | | ECDL (Bauch Welch) | From Emission (Gaussian)    Same as multinomial |
|  | | From transition model |
| Vertibri use DP to find best path. Same recursion except with max not over all nodes  (like,find the max entry of the message rather than suming over the entries (marginalize) the message | | |  |

**Max-product**------find where is the set is the maximum.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  | |  |  |
| In HMM |  |  | | |

**Junction Tree Algorithm------**1.Construct CPT; 2.Convert BN to MRF 3. Triangulate MRF 4. Build Junction Tree 5. probabilities (all polynomial time, NP hard for opt tho)

|  |  |  |
| --- | --- | --- |
| Triangluation= break cycles of >4 nodes | Running intersection property = middle nodes must cotain shared nodes 2 adj. cliques | |
|  |  |  |

First Midterm stuff

|  |
| --- |
| aAW(t - 1)  vv(t)  DC(t)  Classical momentum  .w(t+l)  aAW(t - 1)  Nesterov's  momentum  DC(t') • The hidden activation of the Jth hidden unit the second  hidden la er is the weighted sum of the first hidden layer:  o We can use vector notation to express the hidden vector:  h(2)  10  ZH2  x _ lenMe  30  30 Oe  x) |

Questions

1. Junction Tree step 1
2. EM setting derivative to 0 cant’t give the correction expression for μ