Name: Danny Parker

| Problem | Points | Score |
|---|---|---|
| 1(a) | 15 | |
| 1(b) | 10 | |
| 2(a) | 15 | |
| 2(b) | 10 | |
| 3(a) | 15 | |
| 3(b) | 10 | |
| 4 | 25 | |
| Total | 100 | |

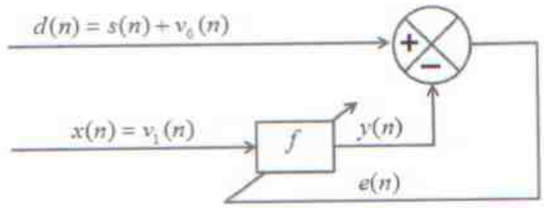
Notes:

(1) The exam is closed books and notes except for two double-sided sheets of notes.

(2) Please indicate clearly your answer to the problem.

(3) The details of your solutions are more important than the answers. Please explain your solutions clearly and include as many details as possible.

**1.** The standard adaptive noise canceller is shown to the right.



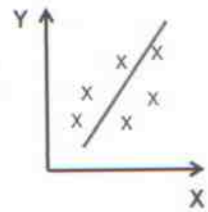

(a) Derive an expression for the optimal filter such that the energy of the noise is minimized.

(b) Suppose instead of this approach you simply computed a linear prediction model on the noisy input signal. Compare and contrast the model you would obtain to the model in (a).

**2.** Recall our expression for a simple linear regression model that formed the basis for maximum likelihood linear regression.

$$Y = \beta_0 + \beta_1 X$$

$$\beta_0 = \overline{Y} - \beta_1 \overline{X} \quad \beta_1 = \frac{\overline{XY} - (\overline{X})(\overline{Y})}{\overline{X^2} - (\overline{X})^2}$$




(a) Derive the optimal value of the slope.

(b) Describe, in qualitative terms, how you would apply this to the problem of adaptation of the model parameters of a Gaussian mixture model. Discuss the pros and cons of this approach.

**3.** In this class we introduced the concept of maximum a posteriori (MAP) adaptation.

(a) Derive the MAP estimate of the mean of a single Gaussian distribution assuming the variance is fixed.

(b) Discuss (but do not derive) properties of this estimate, such as bias. Comment on the implications of the resulting equations in terms of application of this technique to a Gaussian model of common time series such as a speech or image signal.

**4.** In this course, we discussed a range of adaptation topics beginning with the least mean square error (LMS) approaches and ending with approaches based on discriminative training. Describe the course in terms of a tree where the root node is labeled ECE 8423, and all other topics are arranged in a hierarchy representing their relationships with each other. Then provide a glossary: describe the essence of each term represented at each node in a small number of sentences.

Do not feel constrained by the way I presented the course – there is not only one correct answer. Your answers will be judged on their own merits based upon the amount of insight you demonstrate and the completeness of your hierarchy.

ECE 8423 | Final | Danny Parker

1. Derive optimal LMS Filter

goal is to minimize energy of noise. Minimizing the energy of the noise squared is the same as minimizing the energy of the noise so define objective function $J$ as $J = E\{e^2(n)\}$

Since $J$ is a concave function to minimize take derivative and set equal to zero.

$$J = E\{e^2(n)\} = E\{[d(n) - f^T x(n)][d(n) - f^T x(n)]^T\}$$
$$= E\{d(n)d^T(n) - 2f^T x(n) + f^T x(n) x^T(n) f\}$$
$$= \sigma_d^2 - 2f^T g + f^T R f \qquad \text{where } R = E\{xx^T\},\ g = E\{dx\}$$

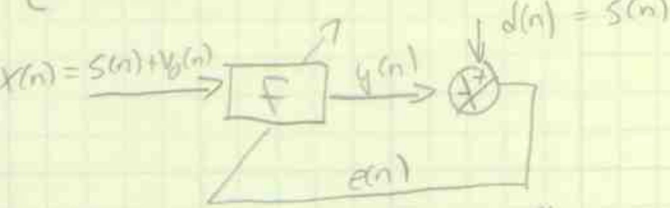
$$\frac{\partial J}{\partial f} = -2g + 2Rf = 0$$
$$Rf = g$$
$$\boxed{f^* = R^{-1} g}$$

for $x(n) = v_1(n)$ iid then $R = \sigma_{v_1}^2 I$

$$g(n) = E\{d(n)x(n)\} = E\{s(n)v_1(n) + v_0(n)v_1(n)\}$$

b) $R = E\{s(n)s(n)\} + \sigma_{v0}^2$

$g = E\{s(n)s(n)\} + \phi$




In part a you have to have access to noise ($v_1$) with similar properties to $v_0$. $v_1$ should have no signal in it. In part b you need access to the signal without noise. In general they operate similar to one another

Danny Parker

2) To minimize take derivative and set equal to zero

$$E\{Y\} = E\{B_0\} + E\{B_1 X\}$$

I'm drawing a blank right now but I'm pretty sure the trick is to substitute one variable in terms of another

$$B_1 = \frac{\overline{XY} - \bar{X}\bar{Y}}{\bar{X^2} - (\bar{X})^2}$$

b) You can use MLLR to reestimate many unknown parameters from a few observations. This is beause you are estimating the transform instead of the parameters directly. The advantages are that it does not require events or states to be directly observed and it can adapt for unobserved states

3) $\underset{\theta}{\operatorname{argmax}}\ p(\theta|x)$

using Bayes Rule $p(\theta|x) = \dfrac{p(x|\theta)\,p(\theta)}{p(x)}$ only need to maximize numerator

$$\underset{m}{\operatorname{argmax}}\ \frac{1}{\sqrt{2\pi\sigma_m^2}} e^{-\frac{1}{2}\left(\frac{m-m_0}{\sigma_m}\right)^2} \frac{1}{\sqrt{2\pi\sigma_v^2}} e^{-\frac{1}{2}\left(\frac{x-m}{\sigma_v}\right)^2}$$

Maximizing $e^{-x}$ is the same as minimizing $x$

$$\hat{m}_{map} = \underset{m}{\operatorname{argmin}} \left[ \frac{1}{2}\left(\frac{m-m_m}{\sigma_m}\right)^2 + \frac{1}{2}\left(\frac{x-m}{\sigma_v}\right)^2 \right]$$

differentiate and set equal to zero $\frac{\partial}{\partial m}$

$$= \frac{1}{2}\left(\frac{1}{\sigma_m^2}\right) 2(m-m_m) + \frac{1}{\sigma_v^2}(x-m)(-1) = 0$$

$$\frac{1}{\sigma_m^2}(m-m_m) - \frac{1}{\sigma_v^2}(x-m) = 0$$

$$m\left(\frac{1}{\sigma_m^2} + \frac{1}{\sigma_v^2}\right) = \frac{x}{\sigma_v^2} + \frac{m_m}{\sigma_m^2}$$

$$m\left(\frac{\sigma_v^2 + \sigma_m^2}{\sigma_m^2\sigma_v^2}\right) = \frac{x}{\sigma_v^2} + \frac{m_m}{\sigma_m^2}$$

$$\hat{m}_{map} = \left(\frac{x}{\sigma_v^2} + \frac{m_m}{\sigma_m^2}\right)\left(\frac{\sigma_m^2\sigma_v^2}{\sigma_v^2+\sigma_m^2}\right) = \boxed{\frac{\sigma_m^2 x + \sigma_v^2 m_m}{\sigma_v^2 + \sigma_m^2}}$$

b) As $N \to \infty$ it will converge to the sample mean of the new data.

for multiple samples

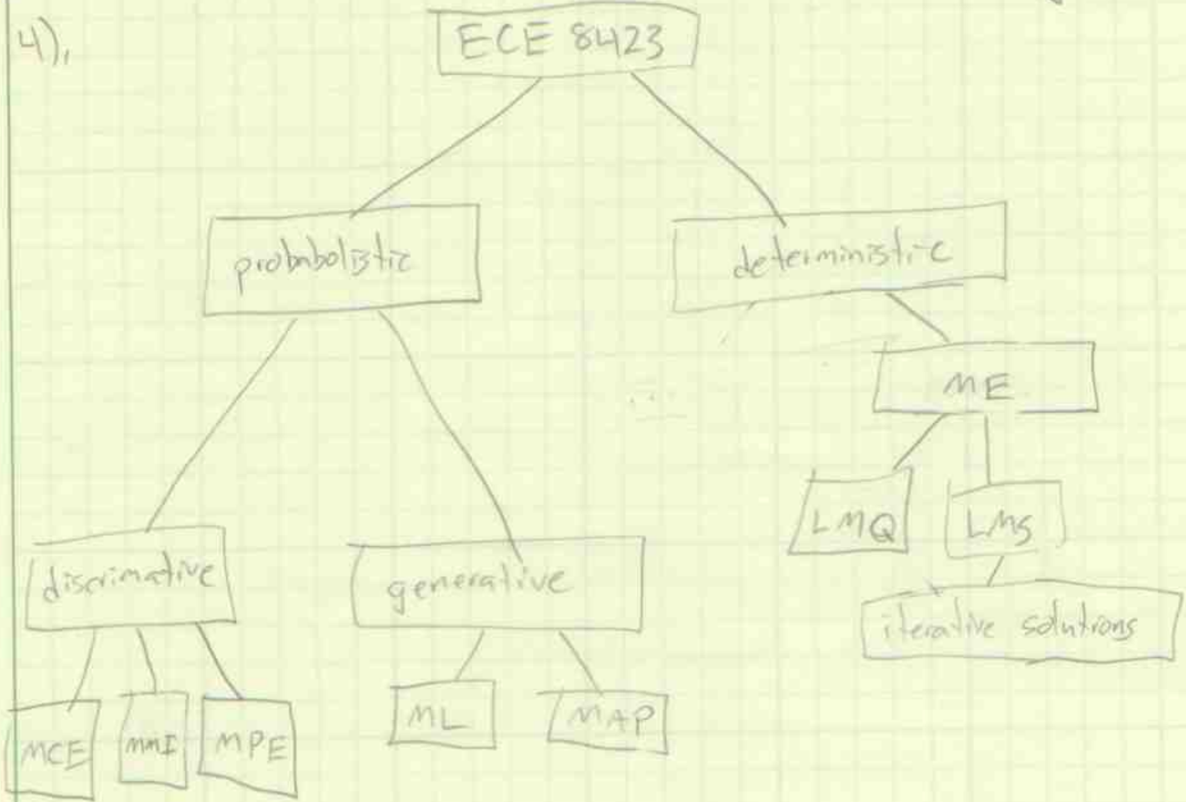
$$\boxed{\frac{\sigma_m^2 \sum^N x_i + \sigma_v^2 m_m}{\sigma_v^2 + N\sigma_m^2}}$$

As $N \to 0$ it converges on the prior mean.

So in effect this mixes the prior mean and sample mean using a weighting that is based on the number of samples and the variances.

Danny Parker

4).




Deterministic - no probabilities

MSE - where the objective is to minimize the mean

LMS - minimize the squared error
LMQ - minimize the fourth error

iterative solutions - various methods to perform the minimization (RLS, Newton, etc) iteratively

Probabolistic - methods that use pdf's

generative - methods that use joint pdf. They maximize prob of correct class.

discrimative - methods that directly estimate class posterior probabilities
They maximize prob of correct class and minimize prob of incorrect one

MCE - Minimum classification error. tries to directly minimize error rate

MMI - Minimizes error rate by removing mutual information between classes

MPE - minimizes "phone" error rates.