Name: Hank Rinehart

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	

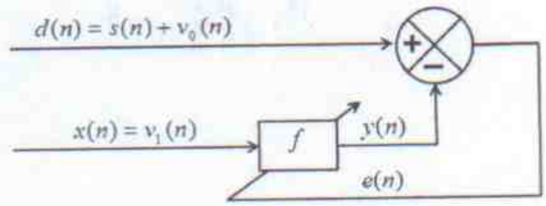
Notes:

(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.



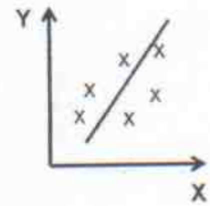
(a) Derive an expression for the optimal filter such that the energy of the noise is minimized.

(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.

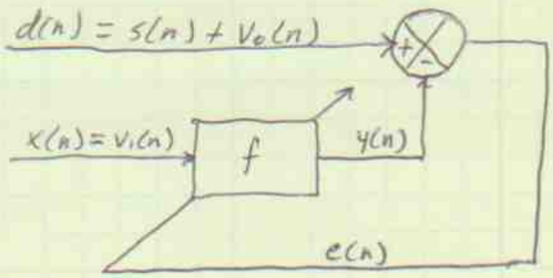
(a) Derive the MAP estimate of the mean of a single Gaussian distribution assuming the variance is fixed.

(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.

1) The standard adaptive noise canceller is shown below:



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

Soln: The error signal can be written as:

$$e(n) = s(n) + v_0(n) - y(n)$$

By taking the expectation of the square of the error:

$$E\{e^2(n)\} = E\{s^2(n)\} + E\{(v_0(n) - y(n))^2\} + 2E\{s(n)(v_0(n) - y(n))\}$$

Assume that $s(n)$ is uncorrelated with $v_0(n)$ and $v_1(n)$

Thus $s(n)$ is uncorrelated with $y(n) = f(n) * v_1(n)$ for a fixed filter

∴ The expectation of the error reduces to:

$$E\{e^2(n)\} = E\{s^2(n)\} + E\{(v_0(n) - y(n))^2\}$$

$E\{e^2(n)\}$ is minimized when $y(n) = v_0(n)$, leaving $s(n) = e(n)$

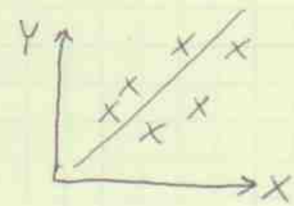
(b) Suppose instead of this approach you simply computed a linear prediction model on the noisy input signal. Compare and contrast the model you would obtain to the model in (a).

Soln: The linear prediction model relies on past samples to develop a prediction as a linear combination of the prior data. The ANC method requires no a priori knowledge of the signal or noise.
ANC requires the use of a secondary reference signal containing little or no signal and consist of noise, correlated in an unknown way, with the noise component of the primary input.

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

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

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



(a) Derive the optimal value of the slope.

Sol'n: Given a sequence of observations $(X_1, Y_1), \ldots, (X_n, Y_n)$

Desire to estimate a function that predicts Y as a function of X

$$y = f(x) = \beta_0 + \beta_1 x$$

Minimize the mean-square error

$$L = \sum_{i=1}^{n} (Y_i - (\beta_0 + \beta_1 X_i))^2$$

minimize wrt β_0 and β_1

Differentiate wrt each parameter

$$\frac{\partial L}{\partial \beta_0} = -\sum_{i=1}^{n} 2(Y_i - (\beta_0 + \beta_1 X_i)) = 0$$

$$\frac{\partial L}{\partial \beta_1} = -\sum_{i=1}^{n} 2(Y_i - (\beta_0 + \beta_1 X_i))X_i = 0$$

Solve for β_0 and β_1:

Let $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$, $\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i$, $\bar{X}^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2$, $\overline{XY} = \frac{1}{n}\sum_{i=1}^{n} X_i Y_i$

$\therefore \; -\sum_{i=1}^{n} 2(Y_i - (\beta_0 + \beta_1 X_i)) = 0 \Rightarrow \bar{Y} = \beta_0 + \beta_1 \bar{X}$

$\& \; -\sum_{i=1}^{n} 2(Y_i - (\beta_0 + \beta_1 X_i))X_i = 0 \Rightarrow \beta_0 \bar{X} + \beta_1 \bar{X}^2 = \overline{XY}$

$$\therefore \; \beta_1 = \frac{\overline{XY} - (\bar{X})(\bar{Y})}{\bar{X}^2 - (\bar{X})^2} \quad \text{and} \quad \beta_0 = \bar{Y} - \beta_1 \bar{X}$$

The optimal value of the slope is:

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

2) (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.

Sol$^{\underline{n}}$: Application:

Pros and Cons: This is a linear regression approach that makes use of a priori information. However ~~this~~ because it also has uniform weighting, outliers will be weighted the same as any other points. Since it is a mean "square" error approach, the farther away the outlier, the more it will impact the fit.

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.

Sol'n: Given $N(\mu_v, \sigma_v)$ distribution with a set of samples $\bar{O} = \{o_1, o_2, \ldots, o_N\}$ drawn for this distribution

Assumptions: 1) μ_v and σ_v were estimated from a large dataset
2) Conditional independence

$$\therefore \hat{\mu}_{MAP} = \underset{\mu}{\operatorname{argmax}}\, p(\bar{O}|\mu)\, g(\mu)$$

$$= \frac{1}{\sqrt{2\pi}\sigma_m} e^{-\frac{1}{2}\left(\frac{\mu-\mu_m}{\sigma_m}\right)^2} \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi}\sigma_v} e^{-\frac{1}{2}\left(\frac{o_i-\mu}{\sigma_v}\right)^2}$$

Maximization of this function translates to minimizing:

$$\hat{\mu}_{MAP} = \underset{\mu}{\operatorname{argmin}} \left[\frac{1}{2}\left(\frac{\mu-\mu_m}{\sigma_m}\right)^2 + \sum_{i=1}^{N} \frac{1}{2}\left(\frac{o_i-\mu}{\sigma_v}\right)^2\right]$$

Differentiate wrt μ:

$$\frac{\partial}{\partial\mu}\left[\frac{1}{2}\left(\frac{\mu-\mu_m}{\sigma_m}\right)^2 + \sum_{i=1}^{N}\frac{1}{2}\left(\frac{o_i-\mu}{\sigma_v}\right)^2\right] =$$

$$= \frac{\mu-\mu_m}{\sigma_m^2} + \left(-\sum_{i=1}^{N} \frac{o_i-\mu}{\sigma_v^2}\right) = 0$$

$$\mu\left(\frac{1}{\sigma_m^2} + \frac{N}{\sigma_v^2}\right) = \frac{\sum_{i=1}^{N} o_i}{\sigma_v^2} + \frac{\mu_m}{\sigma_m^2}$$

$$\mu\left(\frac{\sigma_v^2 + N\sigma_m^2}{\sigma_m^2\sigma_v^2}\right) = \frac{\sum_{i=1}^{N} o_i}{\sigma_v^2} + \frac{\mu_m}{\sigma_m^2}$$

$$\mu = \left(\frac{\sum_{i=1}^{N} o_i}{\sigma_v^2} + \frac{\mu_m}{\sigma_m^2}\right)\left(\frac{\sigma_m^2\sigma_v^2}{\sigma_v^2 + N\sigma_m^2}\right)$$

$$\Rightarrow \hat{\mu}_{MAP} = \frac{\sigma_m^2 \sum_{i=1}^{N} o_i + \sigma_v^2 \mu_m}{\sigma_v^2 + N\sigma_m^2}$$

3) (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.

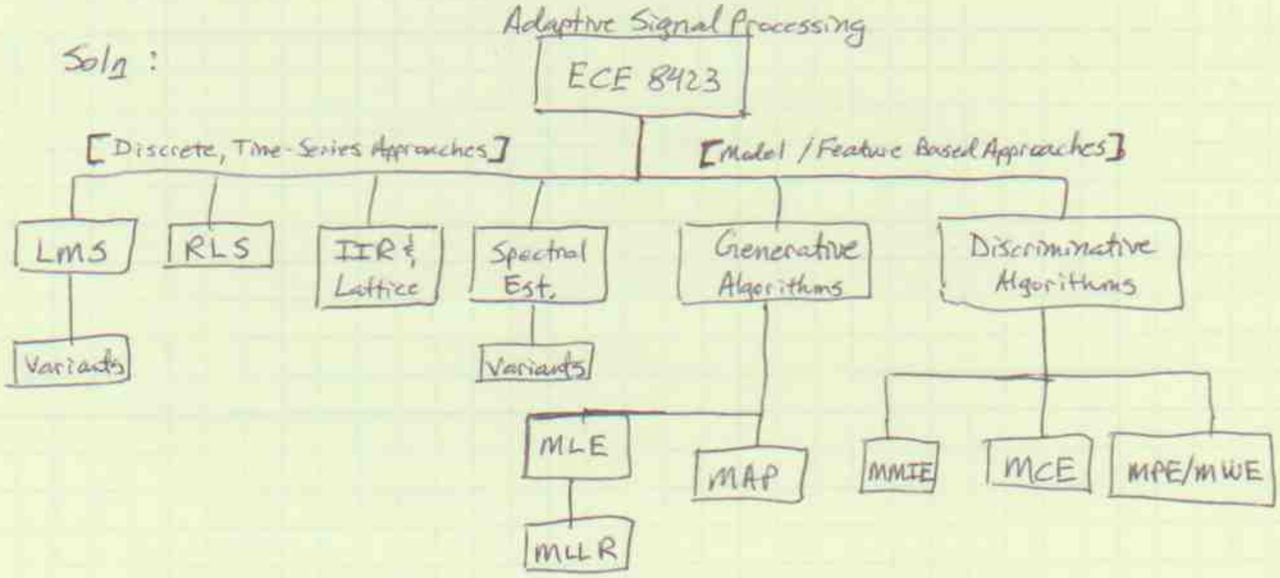
Soln: As the number of samples, N, of new data goes to infinity, the MAP estimate converges to the sample mean.

As N goes to zero, the MAP estimate converges to the prior mean

The adapted mean is a weighted mixture of the sample and prior mean.

4) In this course, we discussed a range of adaptation topics beginning w/ LMS error 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.

Sol$_n$:



Approaches Based on Sampled time Data :

LMS - Least Mean Square : Least-squares minimization of the error
Many variants: Normalized LMS, Leaky, Volterra etc.

RLS - Recursive Least Squares : Goal: Maintains optimal solution at each iteration by using all of the data.

IIR - Infinite Impulse Response

Spectral Estimation : Frequency domain, frame-based approach
Many variants : Periodogram, Blackman-Tukey, etc.

Model and/or Feature Based Approaches

Generative Algorithms : MLE - Maximum Likelihood Estimation
↳ MLLR - Max. Likelihood Linear Regression
MAP - Maximum a posteriori

Discriminitive Algorithms : MMIE - Maximum Mutual Information Estimation
MCE - Minimum Classification Error
MPE/MWE - Minimum Phone/Word Error