They are due Thursday April 27th in class. You do not have to do the problem with extra credit (Problem 5 part 2 and Problem 6).

**Problem 1: MLE and confidence interval.**
Suppose that $X_1, \ldots, X_n$ are i.i.d. uniformly distributed in the interval $[0, \theta]$. Here $\theta$ is the unknown parameter to be estimated from the samples.

1. Find the MLE of $\theta$, denoted by $\hat{\theta}_n$.
2. Prove that $\hat{\theta}_n$ converges to $\theta$ in probability, i.e., for any $\varepsilon > 0$,
   $$\mathbb{P}[|\hat{\theta}_n - \theta| \leq \varepsilon] \to 1$$
as $n \to \infty$.
3. For the error in the MLE, $\theta - \hat{\theta}_n$, compute the probability $\mathbb{P}[n(\theta - \hat{\theta}_n) \geq x]$ for any $n \geq 1$ and $x \geq 0$, and find the limiting distribution of $n(\theta - \hat{\theta}_n)$ as $n \to \infty$.
4. Using the result you obtained above to construct 95% confidence interval in your estimation of $\theta$.

**Problem 2: Laplace distribution.**
Suppose $X_1, \ldots, X_n$ is sampled independently from a double exponential distribution with the density given by
$$f(x|\mu, \sigma) = \frac{1}{2\sigma} \exp\left(-\frac{|x - \mu|}{\sigma}\right), \ -\infty < x < \infty$$

1. Assuming all $X_i \neq X_j$ for all $i \neq j$, can you find the MLE of $\mu$? What if $n$ is odd? What if $n$ is even? Is the MLE unique?
2. Suppose you have determined $\hat{\mu}_n$, find the MLE of $\sigma$.

**Problem 3: Small noise approximation.**
We have a simple dynamical system given by
$$\frac{dx(t)}{dt} = c \ast x(t), \ x(0) = x_0$$
where $c$ is some growth rate and $x_0$ is the initial data. They may have been measured with uncertainties.

1. If the only random source is $x_0$ and it has a distribution $N(\mu, \sigma^2)$, find the distribution of $x(t)$.
2. If the only random source is $c$ and it has a distribution $N(\mu, \sigma^2)$, what is the distribution of $x(t)$? If $\sigma \ll 1$, find the Gaussian approximation to this distribution.
3. We assume that $x_0, c$ are both random and $(x_0, c)$ has joint Gaussian distribution with mean $\mu = (a, b)$ and covariance matrix $\Sigma = (\Sigma_{ij})_{i,j=1,2}$. Can you find the distribution of $x(t)$? If the uncertainties are small, e.g., max$_{i,j=1,2} |\Sigma_{ij}| \ll 1$, find the approximate distribution of $x(t)$.

**Problem 4: MLE of AR(1) process.**
A Gaussian AR(1) process has the form
$$Y_{n+1} = c + \rho Y_n + X_n, \ n = 1, 2, \ldots$$
with $X_n$ i.i.d. $N(0, \sigma^2)$. Here $c, \rho$ are constants and $|\rho| < 1$. 

1. Assuming \( Y_1 \sim N(\mu, \lambda^2) \) and that it is independent of \( \{X_n, n = 1, 2, \ldots\} \), find \( \mu, \lambda^2 \) so that \( Y_n \) has the same distribution for all \( n \in \mathbb{Z}_+ \).

2. Assuming \( Y_1 \) is sampled from the distribution obtained in step 1. Denote by \( \theta = (c, \rho, \sigma^2) \) the parameters to be estimated, find the log-likelihood function of AR(1) model for the sample \( Y_1, \ldots, Y_n \).

3. Assuming now the first observation \( Y_1 = y_1 \) is given (deterministic), find the log-likelihood function of the AR(1) model for the sample \( Y_1, \ldots, Y_n \). Find the MLE of \( \theta \) explicitly.

**Problem 5: MLE in Poisson regression**

Given a set of basis functions \( \{f_k(t)\}_{k=1,\ldots,d-1} \), define

\[
F(t) = \sum_{k=1}^{d-1} \theta_k f_k(t),
\]

and for each \( t \in [0, 1] \), \( Y(t) \) is a Poisson random variable with parameter \( e^{F(t)} \). Given independent samples \( (t_i, Y(t_i))_{i=1,\ldots,N} \), the goal is to use MLE to estimate \( \theta = (\theta_1, \ldots, \theta_{d-1}) \).

1. Formulate your MLE problem by writing down the log-likelihood function.

2. (extra credit: programing exercise) For any \( d \), the basis function is given by

\[
f_k(x) = f \left( \frac{d\pi}{2} (x - \frac{k - 1}{d}) \right), \quad k = 1, \ldots, d - 1,
\]

where

\[
f(x) = \begin{cases} 
\sin(x), & \text{if } x \in [0, \pi], \\
0, & \text{otherwise}.
\end{cases}
\]

Let \( \theta_k = 1 \) for all \( k = 1, \ldots, d - 1 \) be the parameter we want to estimate. Let \( t_i = \frac{i}{N}, i = 1, \ldots, N \).

For \( N = 100, 500, 1000, d = 4, 10, 30, 60 \), do the following:

- generate the data \( (t_i, Y(t_i))_{i=1,\ldots,N} \).
- find \( \hat{\theta} \) by maximizing the log-likelihood function. You may want to use the multidimensional Newton-Raphson method.
- discuss how your results depend on \( N \) and \( d \).

**Problem 6: Linear Discriminant Analysis (extra credit)**

Linear discriminant analysis, also called Gaussian discriminant analysis, is a classic generative learning algorithm in machine learning. In this model, suppose \( X \in \mathbb{R}^d \) is the feature vector, \( Y \in \{0, 1\} \) is the label. Conditioning on \( Y \), \( X \) has Gaussian distribution depending on \( Y = 0 \) or \( 1 \). More precisely, the joint distribution of \( (X, Y) \) follows:

\[
Y \sim \text{Bernoulli}(p) \quad X \mid Y = 0 \sim N(\mu_0, \Sigma) \quad X \mid Y = 1 \sim N(\mu_1, \Sigma)
\]

Here \( p \in (0, 1), \mu_0, \mu_1 \in \mathbb{R}^d, \Sigma \in \mathbb{R}^{d \times d} \) are the parameters of the model. Suppose now we have a dataset of independent samples \( (X^{(i)}, Y^{(i)}), i = 1, \ldots, n \) (note that each \( X^{(i)} \) is a \( d \)-dimensional vector and each \( Y^{(i)} \) is a scalar).

1. Find the log-likelihood function in the model, and compute the MLE of \( \hat{\theta}, \hat{\mu}_0, \hat{\mu}_1, \hat{\Sigma} \).
2. Suppose we have already found the estimation of \( p, \mu_0, \mu_1, \Sigma \). Then we can use our model to make a prediction given some query point \( X \). Show by Bayes formula that the posterior distribution of the label \( Y \) at \( X \) takes the form of a logistic function, i.e.,

\[
P(Y = 1 | X; p, \mu_0, \mu_1, \Sigma) = \frac{1}{1 + \exp(-\theta^T \tilde{X})}
\]

where \( \theta \) is some appropriate function of \( p, \mu_0, \mu_1, \Sigma \) and \( \tilde{X} = [1; X] \in \mathbb{R}^{d+1} \).