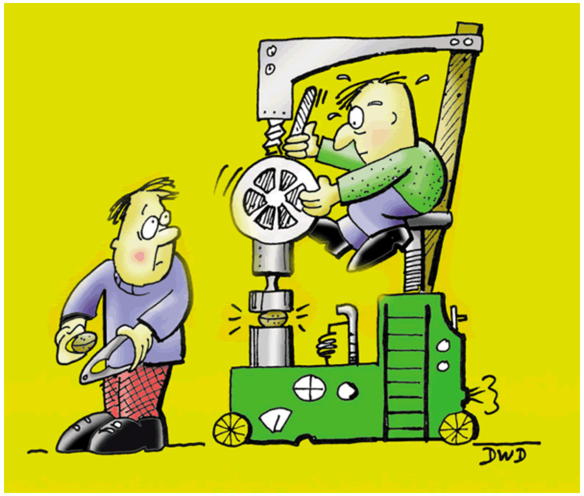
Lecture 25
Model Order Estimation
ECEN 5283 Computer Vision

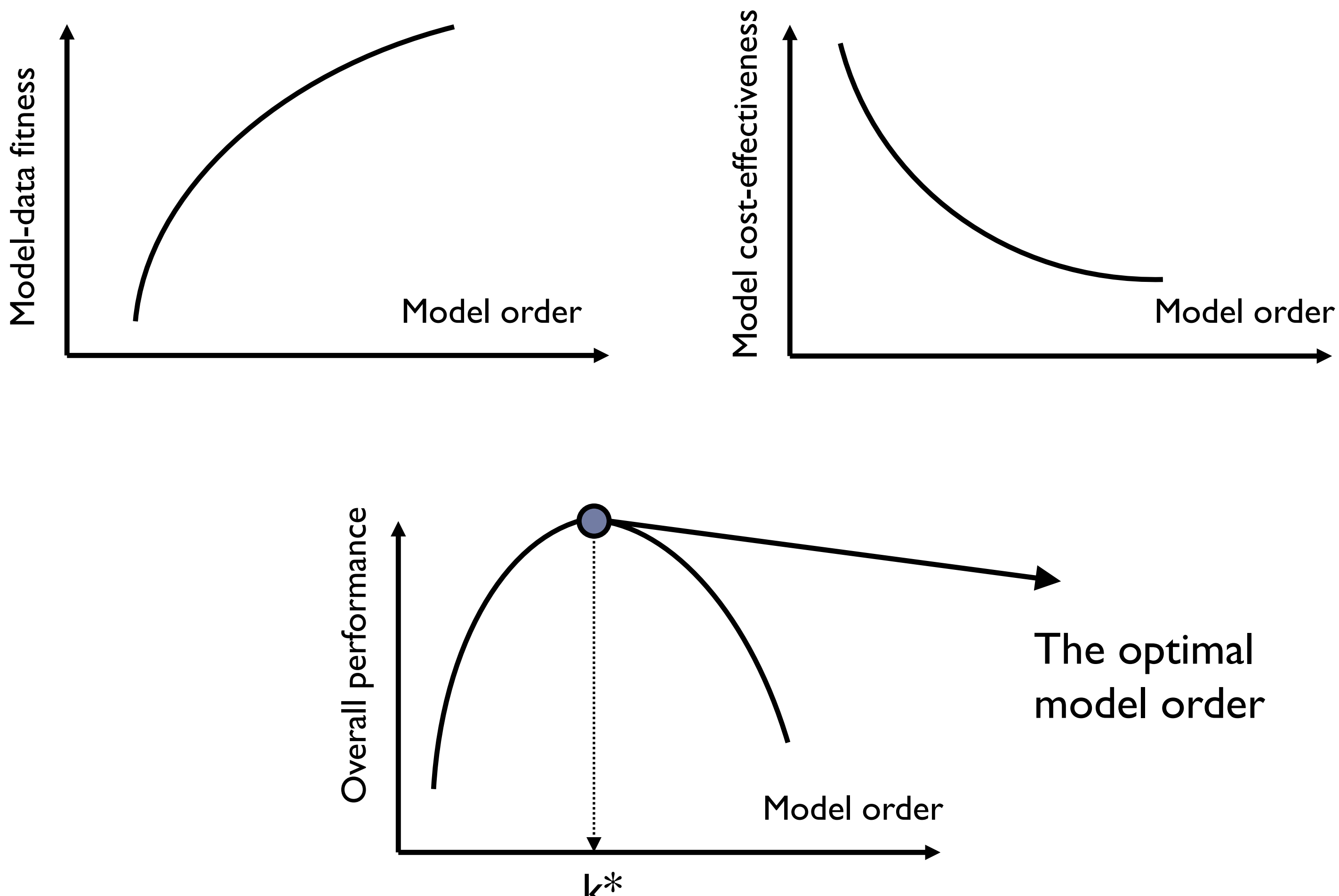
Dr. Guoliang Fan
School of Electrical and Computer Engineering
Oklahoma State University

Goals

- To introduce the model order estimation problem.
- To discuss the minimum description length (MDL) criterion for model order estimation.
- To apply a prior probability for order model estimation.



Basic idea



Model-data fitness
Model order
Model cost-effectiveness
Model order
Overall performance
Model order
The optimal model order
k*

What is the objective function of the EM algorithm?

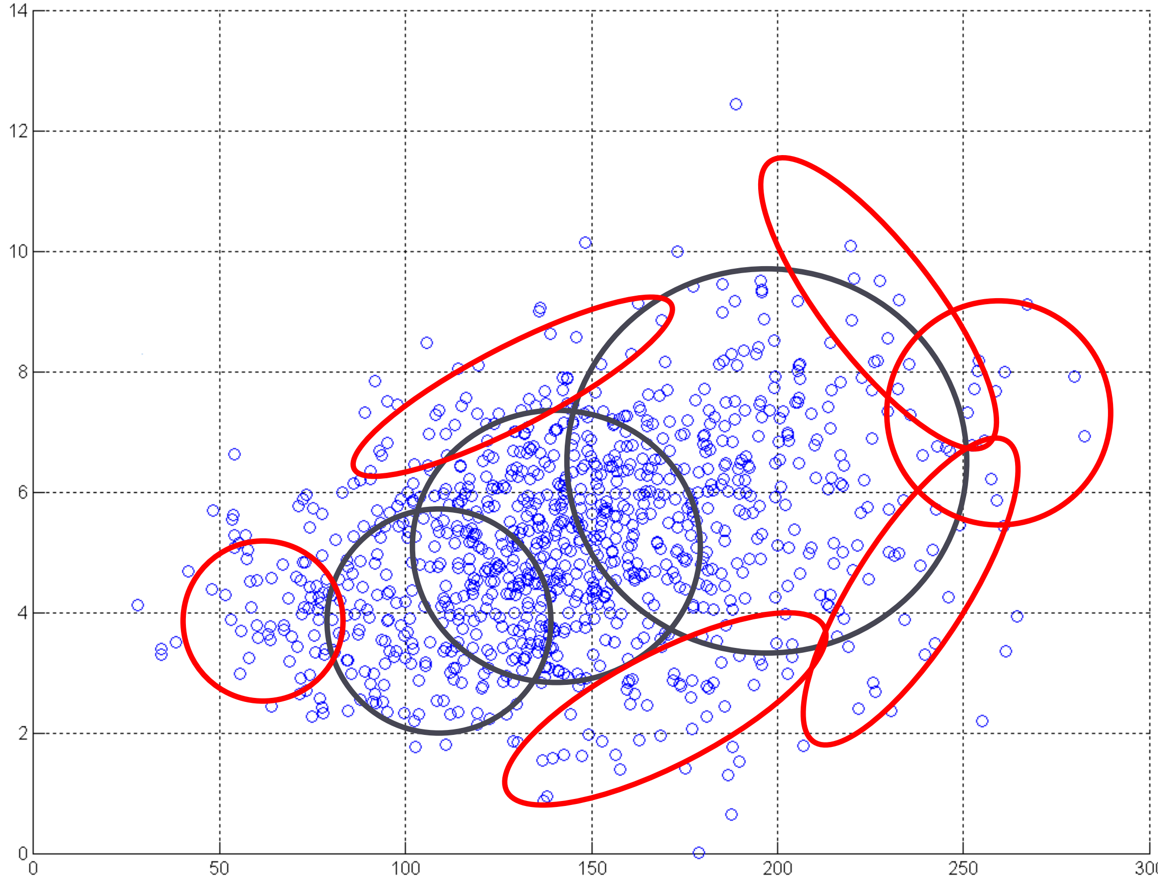
- The objective function of the EM algorithm is the incomplete data log-likelihood defined as

$$\log p(Y \mid \Theta) = \sum_{j=1}^{N} \log\left(\sum_{i=1}^{K} p(y_j \mid x_j = i, \Theta)\alpha_i \right)$$

- If we keep the same objective function for the case that the number of cluster K is unknown as, then we get

$$\log p(Y \mid K, \Theta) = \sum_{j=1}^{N} \log\left(\sum_{i=1}^{K} p(y_j \mid x_j = i, \Theta)\alpha_i \right)$$

Can we use this objective function to estimate the model order?



Can we use this objective function to estimate the model order?

- Can we use the maximum likelihood (ML) estimation to find the model order and parameters as

$$\left(\Theta^*, K^*\right) = \arg\max_{\Theta} \log p(Y \mid K, \Theta)$$

- Unfortunately, the ML estimation of the model order and parameters is not well defined here.
  - A larger number K can always make the likelihood larger (or better fitness with a higher order model).
  - There should be a penalty term to penalize a larger order.

Minimum Description Length (MDL)

- The MDL estimator attempts to find the model order which minimizes the number of bits that would be required to code both the data samples Y and the parameter vector

$$\Theta = (\alpha_1, \ldots, \alpha_K, \mu_1, \ldots, \mu_K, \Sigma_1, \ldots, \Sigma_K,)$$

$$MDL(K, \Theta) = -\log p(Y \mid K, \Theta) + \frac{1}{2} L \log(NM)$$

L : the number of parameters in Θ

M : the dimension of the observation y

N : the size of data samples

Where is K?

https://engineering.purdue.edu/~bouman/software/cluster/manual.pdf

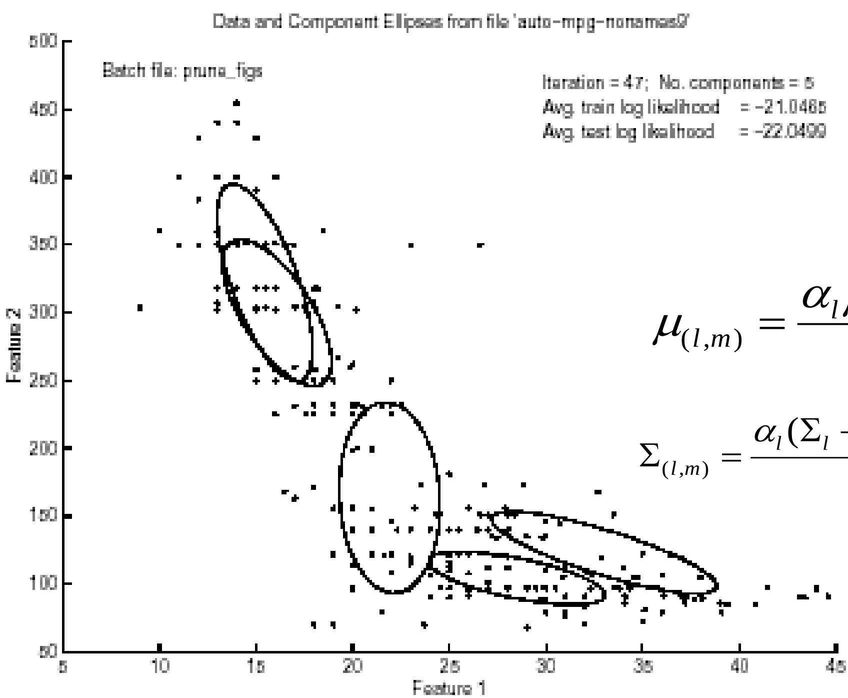
EM algorithm for MDL optimization

- Step 1: Initialize the class number with the largest one $k=K_{max}$.
- Step 2: Initialize the parameters of k Gaussians.
- Step 3: Apply the EM algorithm until MDL(k, Θ) converges.
- Step 4: Record the corresponding Θ* and MDL value.
- Step 5: If k>1, reduce the number of clusters, k=k-1 and go back to step 2.
- Step 6: Choose the value k* and parameter Θ* which minimize the value of MDL.

$$\left(k^*, \Theta^*\right) = \arg_{k=1,\dots,K} \min MDL(k, \Theta)$$

How to trim down the number of clusters?

- In order to provide a good initialization for the lower-order model, we just need to merge to the two Gaussian components which are the closest.



http://cobweb.ecn.purdue.edu/~bouman/software/cluster/

How to compute the distance between two Gaussian components?

$$\alpha_l, \theta_l = \{\mu_l, \Sigma_l\} \qquad \alpha_m, \theta_m = \{\mu_m, \Sigma_m\}$$

The distribution of the composite component if we merge the two

$$\mu_{(l,m)} = \frac{\alpha_l \mu_l + \alpha_m \mu_m}{\alpha_l + \alpha_m}$$

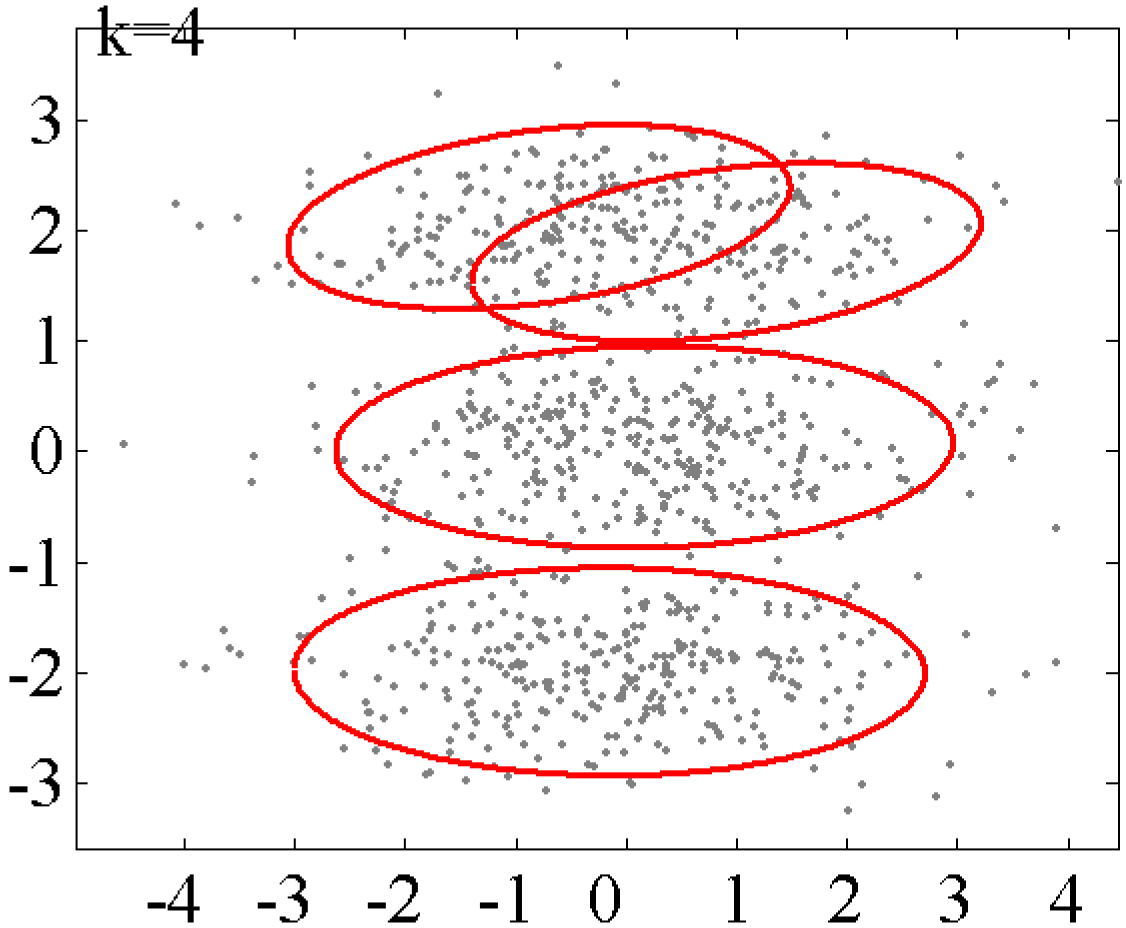
$$\Sigma_{(l,m)} = \frac{\alpha_l(\Sigma_l + (\mu_l - \mu_{(l,m)})(\mu_l - \mu_{(l,m)})^T) + \alpha_m(\Sigma_m + (\mu_m - \mu_{(l,m)})(\mu_m - \mu_{(l,m)})^T)}{\alpha_l + \alpha_m}$$

Distance between two Gaussian components

$$d(l,m) = \frac{N\alpha_l}{2}\log\left(\frac{|\Sigma_{(l,m)}|}{|\Sigma_l|}\right) + \frac{N\alpha_m}{2}\log\left(\frac{|\Sigma_{(l,m)}|}{|\Sigma_m|}\right)$$

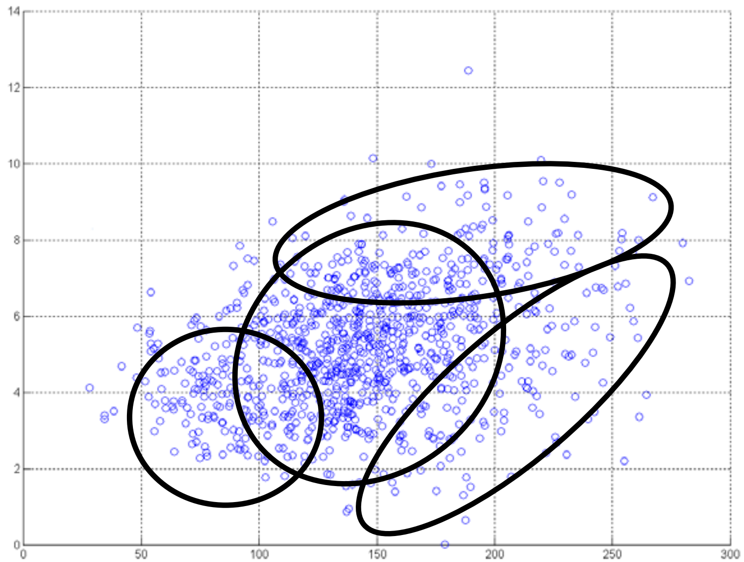
It is the summation of the similarity between each individual component and the composite one.

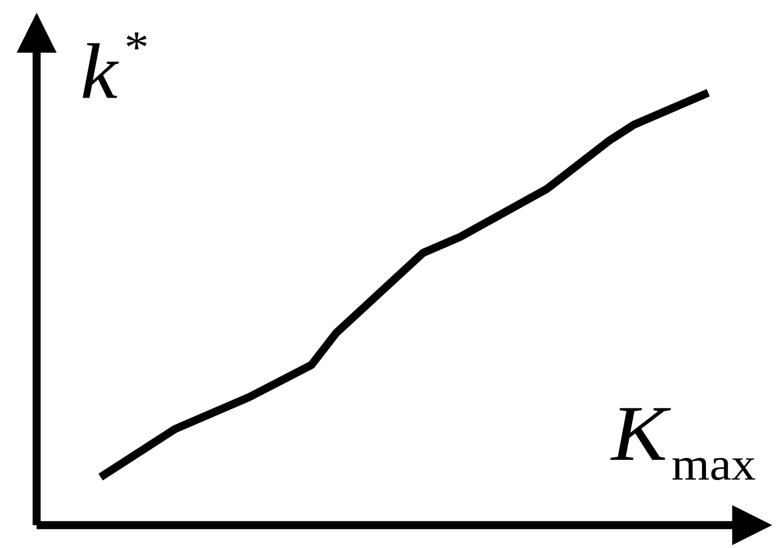
A MDL Example



Shortcoming of MDL Estimation

- There are two major limitations of MDL estimation
  - The optimal model order may not be semantically meaningful. That means the estimated cluster number may not necessary reveal the underlying structure in the data.
  - The optimal model order estimated could depend on K_{max} that is the largest model order.

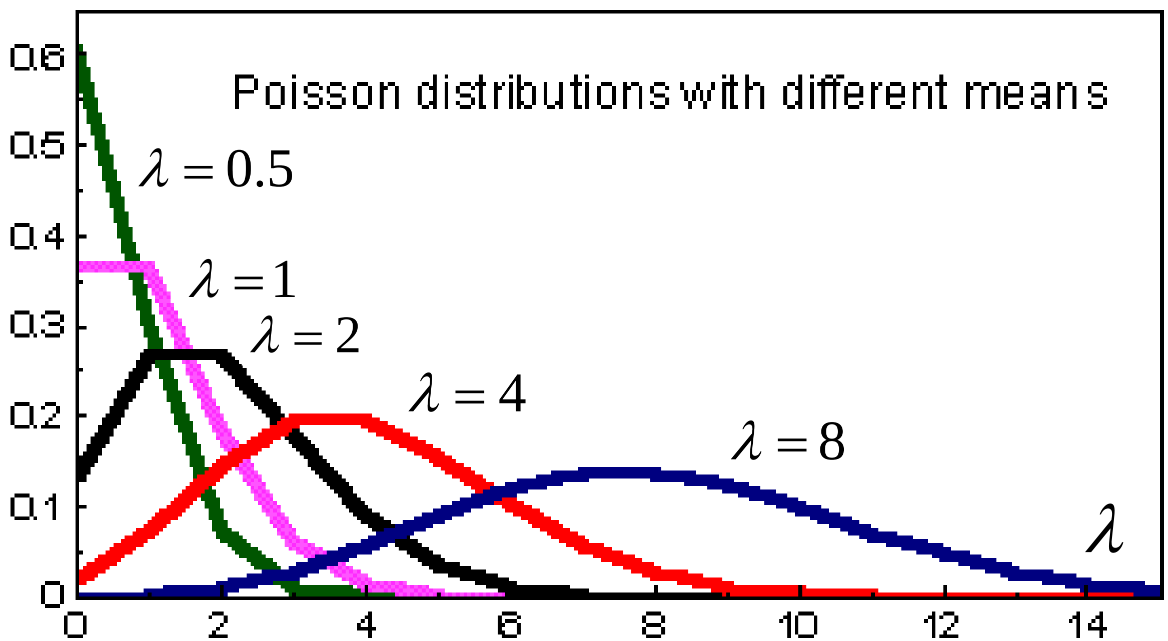




Another Alternative Approach for Model Order Estimation

- If we assume a prior probability of models of different orders as a Poisson distribution as

$$P(K^* = k) = \frac{\lambda^k}{e^\lambda k!} \qquad \begin{array}{l}\text{mean } \lambda \\ \text{variance } \lambda\end{array}$$



A New Objective Function

- Given a prior probability of the model Θ_k, we can define a new joint posterior probability density to represent the solution

$$p(\Theta, k \mid Y) = p(\Theta \mid Y)\, p(k \mid Y) = p(\Theta \mid Y)\, p(k)$$

What assumption is used here?

What assumption is used here?

$$p(\Theta \mid Y) = \frac{p(\Theta, Y)}{p(Y)} = \frac{p(Y \mid \Theta)\, p(\Theta)}{p(Y)}$$

$$p(\Theta, k \mid Y) \propto \underbrace{p(Y \mid \Theta)}_{\text{Data likelihood}}\, \underbrace{p(k)}_{\text{Model prior}}$$

Optimization Algorithm

- Step 1: Initialize k=1;
- Step 2: Optimize the model parameter Θ_k using EM;
- Step 3: Compute the data likelihood $p(Y|\Theta_k)$ and the prior probability of model order k, i.e., $P(k)$;
- Step 4: $k=k+1$ and go to step 2, until $k=K_{max}$;
- Find the model order that has be the largest joint posterior probability.

$$\{k^*, \Theta^*\} = \arg_{\{k,\Theta\}} \max p(\Theta, k \mid Y)$$
$$= \arg_{\{k,\Theta\}} \max p(Y \mid \Theta)\, p(k)$$
$$= \arg_{\{k,\Theta\}} \max\left(\log p(Y \mid \Theta) + \log p(k)\right)$$

The Main Challenge

- The prior probability has to be carefully set in order to produce a meaningful result on model order estimation.
  - In practice, it is more desirable to set λ relative smaller. Why?