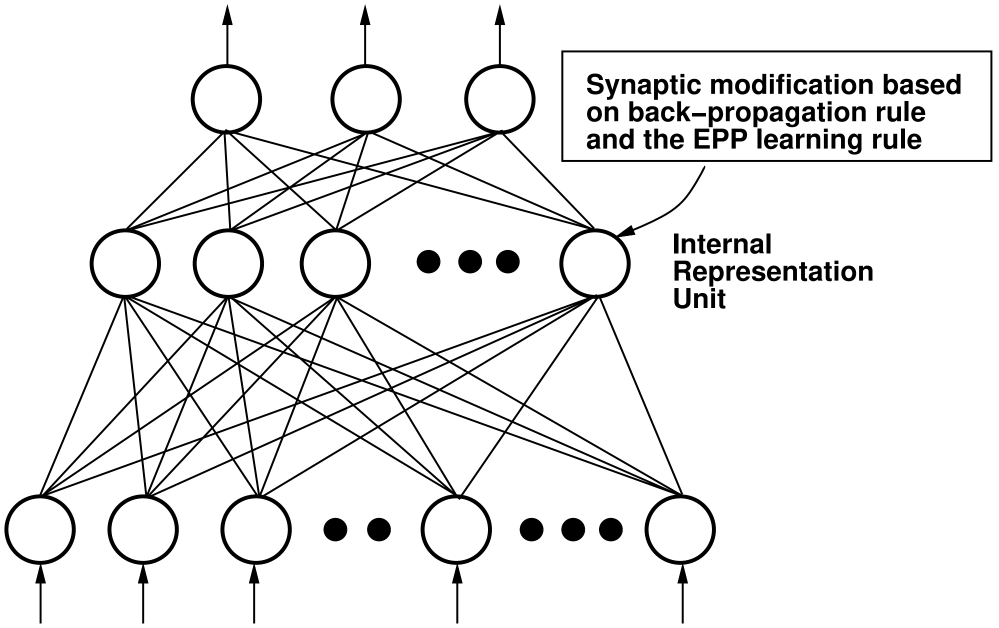
A Hybrid Modeling Approach

- When the model is unknown or the ratio between the number of training samples and their dimensionality is low:
  - A feed-forward architecture is very useful in incorporating prior knowledge and varius model constraints
- The learning rule is the gradient of two tasks: maximizing class probabilities and maximizing a functional that measures the quality of the hidden units (model constraints)
- Possible measures are: BCM, Kurtosis, Component independence, Entropy.
- The general framework for the unsupervised part is Exploratory Projection Pursuit (EPP) which measures different forms of deviation from a Gaussian Distribution.

The Hybrid learning rule

A hybrid EPP/PPR neural network (EPPNN)
(Intrator, 1993)



- The unsupervised learning rule is applied to the hidden units only.

- The unsupervised learning rule is a means to impose prior knowledge (BIAS) onto the classification scheme and is most useful with high dimensional data and small training set.

$$\frac{\partial w_{ij}}{\partial t} = -\epsilon \, [\frac{\partial \mathcal{E}(w,x)}{\partial w_{ij}} + \frac{\partial \rho(w_1,\ldots,w_n)}{\partial w_{ij}}$$
$$+(\text{Contribution of Cost/Complexity terms})].$$

A Hybrid Reconstruction-Classification Appraoch

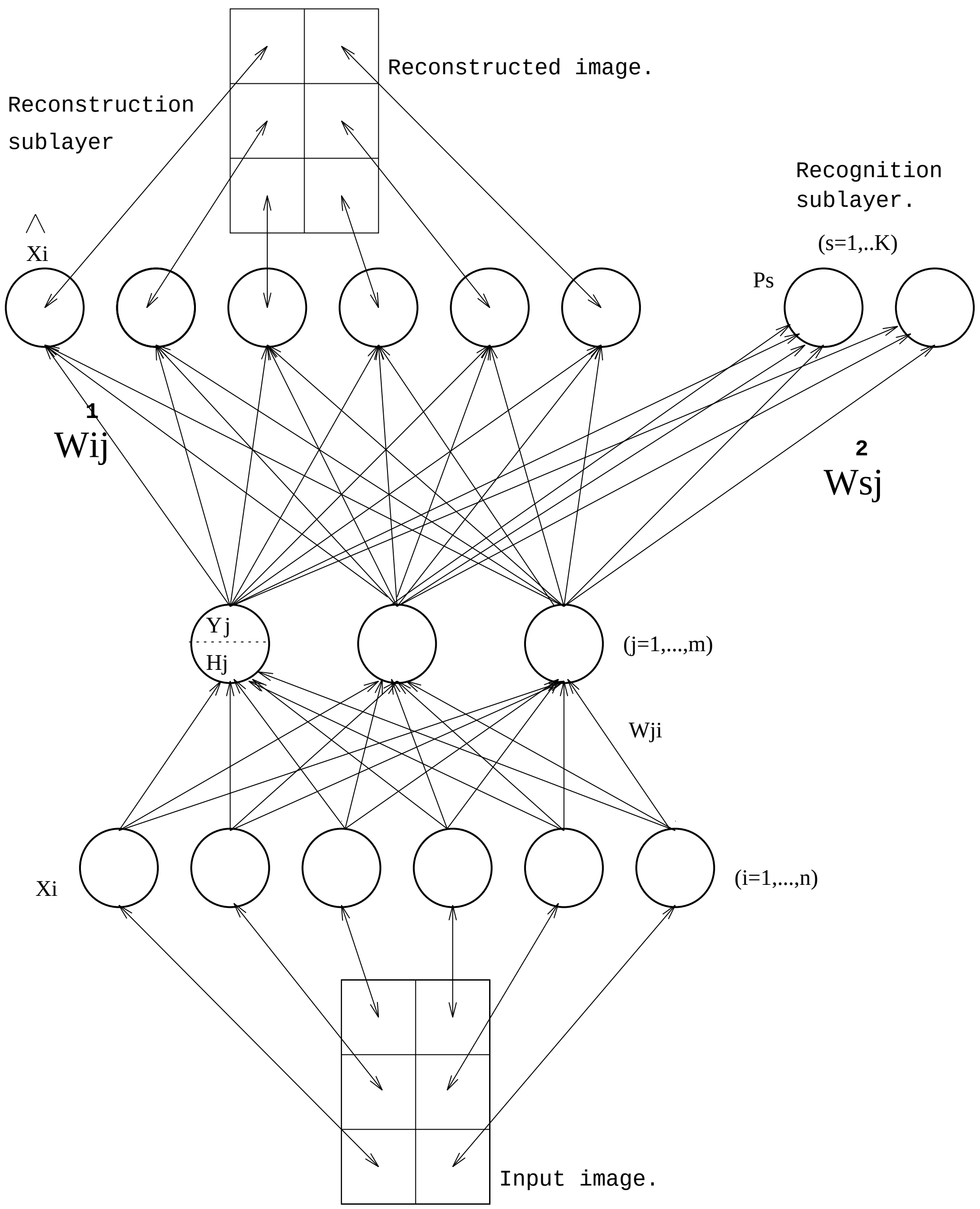
- An important aspect of a model is to provide confidence about its prediction

- One way to get prediction confidence is by modeling concurrently the input data and the desired classification

- This leads to a model that attempts to reconstruct the input data in addition to providing classification

- In general, this also improves model quality

A Reconstruction-constraints Architecture



Motivation for Integration of Experts

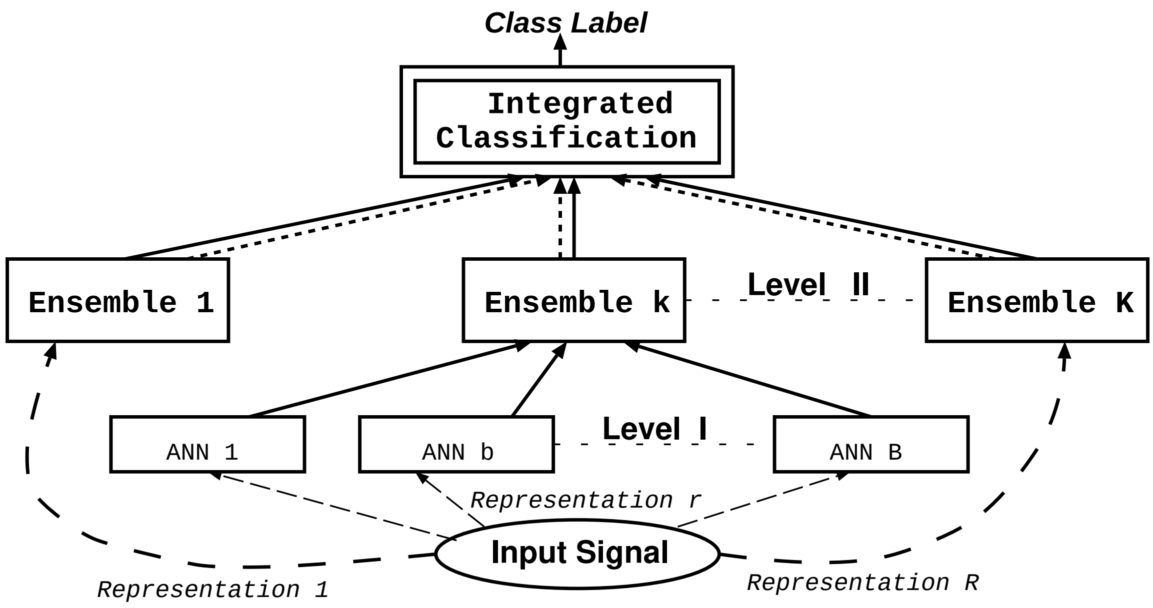
- High-dimensional problem; Local characteristics.
- Non-stationary (Phases) or non-homogeneous domain.
- Noise, incomplete domain knowledge.
- Sparse distribution.

Methods of integration

- Simple integration: Averaging over all ensembles.
- Maximum entropy considerations:
  When the variance of ensemble members is a good estimator for the confidence: the weight of ensemble k should be

$$\text{exp}(-V[E_k])/\sum_i exp(-V[E_i])$$

Integration of Experts



- Each ensemble in Level II recieves the prediction output of a collection of predictors from Level I.

- These Level I predictors receive the same input data representation

- Level I predictors connected to different Level II experts recieve a different input data representation