

# T-61.5030 Advanced Course in Neural Computing Examination 18.12.2006

Karhunen

1. Answer briefly (using a few lines) to the following questions:
  - (a) What VC dimension describes and measures?
  - (b) What for is Oja's rule used in neural computing?
  - (c) What measures Kullback-Leibler divergence?
  - (d) When a stochastic process forms a Markov chain?
  - (e) Which two broad algorithm types are used in dynamic programming for minimizing the cost-and-go function?
  - (f) How second-order neuron differs from usual neuron?
  
2. The figure (on the reverse side) shows a schematic diagram of the mixture-of-experts neural network. For this network structure, there is a corresponding generative probabilistic model for the scalar output  $y$  of the network as a function of its input vector  $\mathbf{x}$  and parameters of the network. Explain:
  - (a) Which kind of mathematical model each expert forms between the input vector and its own output?
  - (b) What is the mathematical model to the total output  $y$  of the network?
  - (c) How the gating network models the mixing coefficients  $g_i$ ?
  - (d) What type of generative probabilistic model the whole network forms between its output  $y$  and input vector  $\mathbf{x}$ ?
  - (e) What properties this mathematical model has?
  - (f) How one can expand the network further to an even more general model?

Please note that you need not discuss learning of the mixture-of-experts network, but only the mathematical model that it provides.

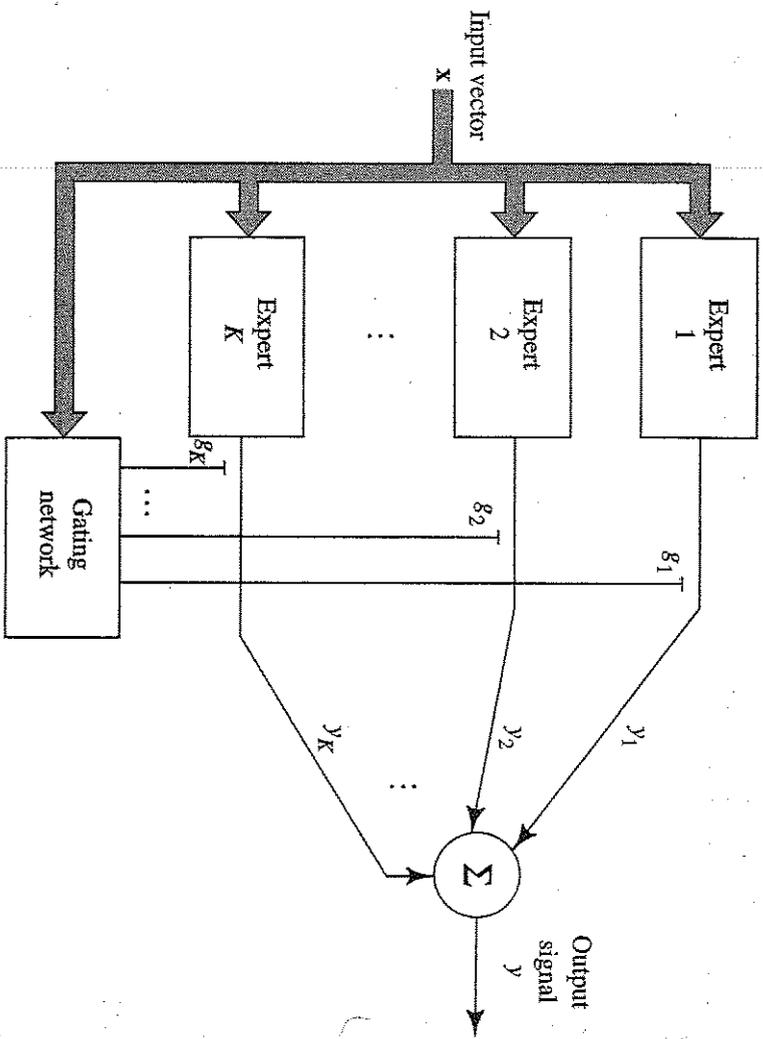
3. Consider a zero mean data vector  $\mathbf{x}$  and a weight vector  $\mathbf{w}$  with the same dimension, normalized to have unit length:  $\|\mathbf{w}\|^2 = 1$ . Then the variance of the projection  $\mathbf{x}^T \mathbf{w}$  of  $\mathbf{x}$  onto  $\mathbf{w}$  is  $E[(\mathbf{x}^T \mathbf{w})^2]$ , where  $E$  is the expectation operator. Find the direction  $\mathbf{w}$  which maximizes this variance. What is the maximum value of the variance?

4. Consider  $n$  zero mean source signals  $s_i$ , which have been collected to the vector  $\mathbf{s} = (s_1, s_2, \dots, s_n)^T$ . Assume that the joint distribution of the sources is  $n$  variables multivariate Gaussian distribution

$$p(\mathbf{s}) = (2\pi)^{-n/2} (\det \mathbf{C})^{-1/2} \exp \left\{ -\frac{1}{2} \mathbf{s}^T \mathbf{C}^{-1} \mathbf{s} \right\}$$

Here  $\mathbf{C} = E\{\mathbf{s}\mathbf{s}^T\}$  is the covariance matrix of the vector  $\mathbf{s}$  and  $\det$  means determinant. Assume that the source signals  $s_1, s_2, \dots, s_n$  are mutually uncorrelated. Show that they are mutually statistically independent.

Does this property hold for other than Gaussian distributed source signals? How about the reverse property: are statistically independent source signals source signals mutually uncorrelated?



- A SCHEMATIC DIAGRAM OF MIXTURE-OF-EXPERTS NETWORK.

- KANONIKOVA 'ASIAUTOTIJDIDEN SEKOTIUS' NEUROVERKOSTA,