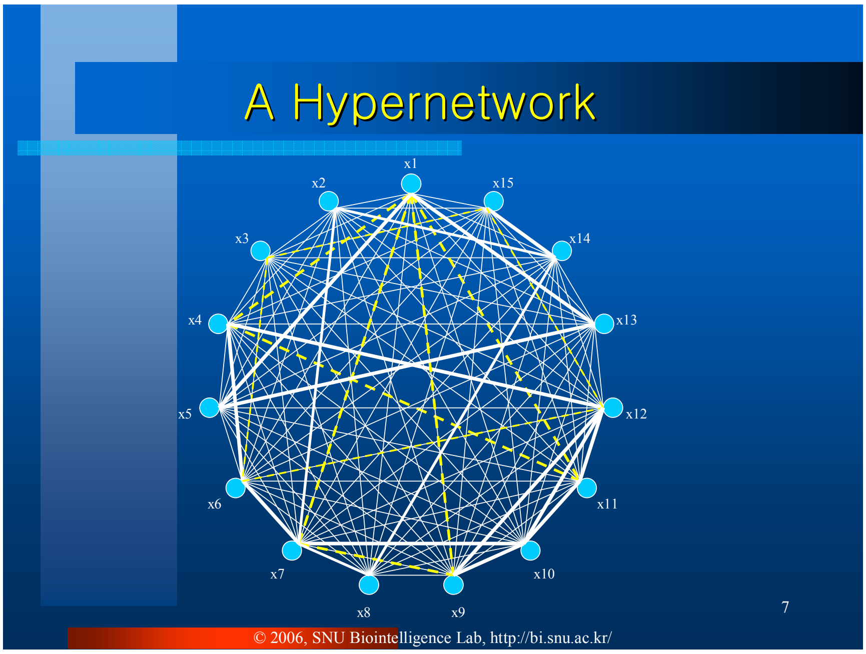
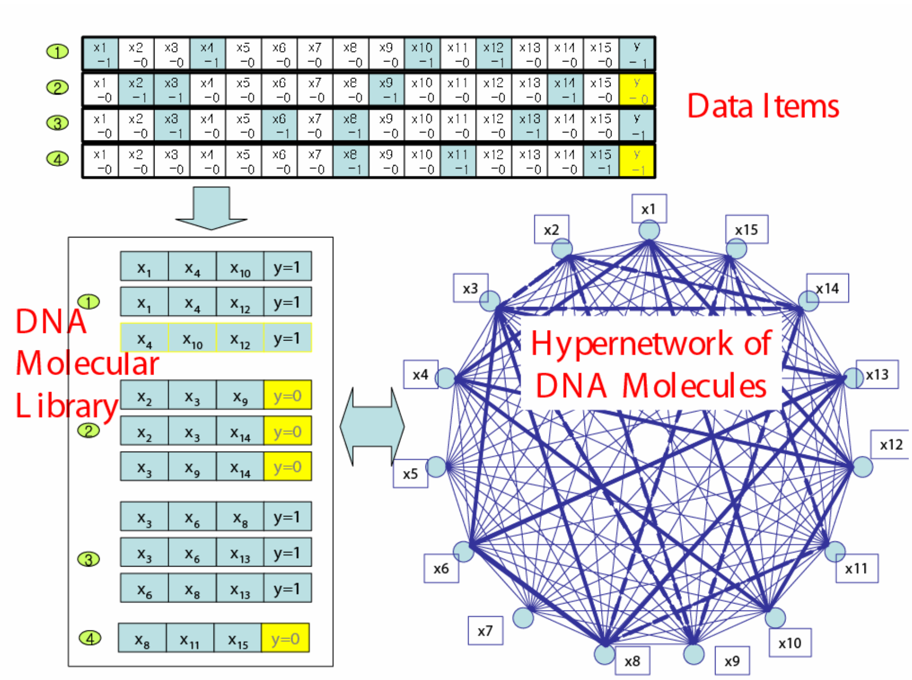
4.2 Hypernetwork Memory

하이퍼망은 가중치를 가진 에지를 포함하는 하이퍼그래프로 정의된다. 즉 하이퍼망은 $H = (X, E, W)$로 표시하며 여기서 X는 정점의 집합이고, E는 에지의 집합, W는 에지에 대응되는 가중치들의 집합이다. 하이퍼망은 정보저장구조 즉 메모리로 사용될 수 있다.





4.2.1 Recall (Completion) Memory Models

하이퍼망은 다음과 같이 확률변수들 $X=(X_1,X_1,...,X_n)$ 간의 결합확률분포(joint probability)를 나타내는데 사용될 수 있다.

$$P(x)=\frac{1}{Z}\prod_{i=1}^{|E|} w_i \varphi_{E_i}(x)$$

$$Z=Z(E,W)=\sum_{x'}\prod_{i=1}^{|E|} w_i \varphi_{E_i}(x')$$

$$E_i=\{x_{i_1},x_{i_2},...,x_{i_{|E_i|}}\} \quad \text{: hyperedge in } H$$

$$\varphi=\varphi_{E_i}: A_{E_i}\rightarrow \Re \quad \text{: potential function}$$

보통 다음과 같은 Gibbs 분포를 사용하며

$$P(x)=\frac{1}{Z}\exp\left(-\sum_{i=1}^{|E|} w_i \varphi_{E_i}(x)\right)=\frac{\exp\left(-\sum_{i=1}^{|E|} w_i \varphi_{E_i}(x)\right)}{\sum_{x'}\exp\left(-\sum_{i=1}^{|E|} w_i \varphi_{E_i}(x')\right)}$$

여기서 $\exp\left(-\sum_{i=1}^{|E|} w_i \varphi_{E_i}(x)\right)$는 Boltzmann factor, $\sum_{i=1}^{|E|} w_i \varphi_{E_i}(x)$는 하이퍼망 시스템의 전체 에너지를 나타내고, $\varphi_{E_i}(x)$는 potentional function이다. $\varphi_{E_i}(x)$는 모델링 하고자 하는 구체적인 대상에 따라서 다음과 같은 형태를 취할 수 있다.

$$\varphi_{E_i}(x)=\prod_{x_j\in E_i} x_j = x_{i_1}x_{i_2}\cdots x_{i_{|E_i|}} \quad \text{: product (conjunction)} \quad \text{(i)}$$

$$\varphi_{E_i}(x)=\sum_{x_j\in E_i} x_j = x_{i_1}+x_{i_2}+\cdots+x_{i_{|E_i|}} \quad \text{: sum (disjunction)} \quad \text{(ii)}$$

$$\varphi_{E_i}(x)=\frac{1}{1+\exp\left(-\sum_{x_j\in E_i} w_j x_j\right)} \quad \text{: sigmoid} \quad \text{(iii)}$$

$$\varphi_{E_i}(x)=\exp\left(-\frac{\| x^{(i)}-x \|_{E_i}}{\sigma_i}\right) \quad \text{: gaussian} \quad \text{(iv)}$$

$$\varphi_{E_i}(x)=\frac{\exp\left(-\sum_{x_j\in E_i} w_j x_j\right)}{\sum_{k=1}^{|E|}\exp\left(-\sum_{x_j\in E_k} w_j x_j\right)} \quad \text{: probability} \quad \text{(v)}$$

이러한 선택성은 메모리와 학습 현상을 유연하게 모델링 하는데 유용하다. 예를 들어, (1.9)와 (1.10)의 (ii)를 조합할 경우 다음과 같은 확률분포를 표현하는 메모리를 구성할 수 있다.

$$
\begin{aligned}
P(x) &= \frac{1}{Z}\exp\left(-\sum_{i=1}^{|E|} w_i \varphi_{E_i}(x)\right) \\
&= \frac{1}{Z}\exp\left(-\sum_{i=1}^{|E|} w_i \prod_{x_j \in E_i} x_j\right) \\
&= \frac{1}{Z}\exp\left(-\sum_{i=1}^{|E|} w_i x_{i_1} x_{i_2} \cdots x_{i_{|E_i|}}\right) \\
&= \frac{1}{Z}\exp\left(-\sum_{k=1}^{n} w_i^{(k)} x_{i_1} x_{i_2} \cdots x_{i_{|E_i^{(k)}|}}\right)
\end{aligned}
$$

$$
Z = \sum_{x'} \exp\left(-\sum_{k=1}^{n} w_i^{(k)} x_{i_1} x_{i_2} \cdots x_{i_{|E_i^{(k)}|}}\right)
$$

위에서 $E_i^{(k)}$ 는 에지 E_i 의 cardinality가 k임을 명시적으로 나타낸다. 마찬가지로 $w_i^{(k)}$ 도 cardinality가 k인 에지의 가중치임을 표시한다

이와 같이 하이퍼망 H = (X, E, W)가 메모리를 표현하고 있으며 특히 하이퍼에지 E가 빈도수 W가 구체적인 메모리 요소들의 분포를 나타내고 있다. H에서 cardinality k인 에지만으로 구성된 $E^{(k)}$를 생각하자. 패밀리 $E^{(k)} \subset E$ 로 구성된 일부 하이퍼망을 H의 **일부 k-하이퍼망(partial k-hypernetwork)**이라 하고 이를 $H^{(k)}$로 나타낸다.

$$
\begin{aligned}
H^{(k)} &= (X^{(k)}, E^{(k)}, W^{(k)}) \\
E^{(k)} &= \{E_i \mid E_i \in E;\ |E_i| = k\} \subset E \\
X^{(k)} &= \bigcup_{E_i \in E^{(k)}} E_i \\
W^{(k)} &= \{W_i \mid E_i \in E^{(k)}\}
\end{aligned}
$$

H에 대한 일부 하이퍼망임을 문맥으로부터 알 수 있는 경우에는 $H^{(k)}$를 k-하이퍼망이라고도 부르기로 한다. 일반적인 용어로서 k-하이퍼망은 모든 에지의 cardinality가 k로 균일한 하이퍼망 즉 **k-균일 하이퍼망**을 뜻한다.

또한 여러 개의 k 값에 대한 k-하이퍼망들의 집합을 $(k_1,k_2,...,k_m)$ **섹션**으로 정의하고 다음과 같이 $H^{(k_1,k_2,...,k_m)}$ 로 표시한다.

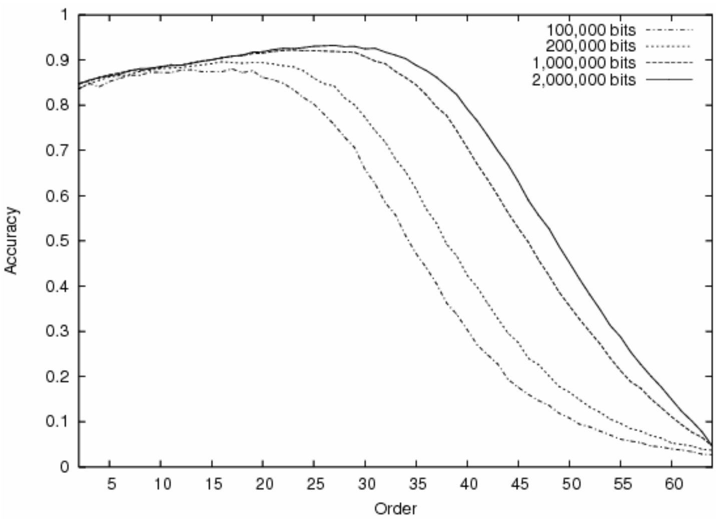
$$
H^{(k_1,k_2,...,k_m)} = H^{(k_1)} \cup H^{(k_2)} \cup \cdots \cup H^{(k_m)}
$$

연속된 구간의 k 값에 대한 k-하이퍼망들의 집합은 특별히 $H^{(k_1 \dots k_2)}$ 로 표시하고 $(k_1...k_2)$ **섹션**이라 한다.

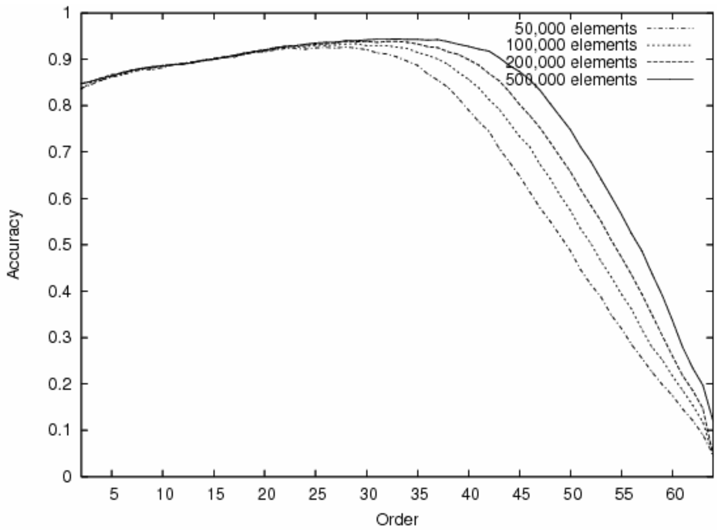
$$
H^{(k_1 \dots k_2)} = H^{(k_1)} \cup H^{(k_1+1)} \cup H^{(k_1+2)} \cup \cdots \cup H^{(k_2)}
$$

하이퍼망 H의 $(k_1...k_2)$-섹션에 대한 가중치집합 $W^{(k_1 \dots k_2)}$ 로 나타내기로 하자. Cardinality 값 k=k_1,k_1+1,...,k_2에 대한 하이퍼에지의 빈도수 즉 $|W^{(k)}|$의 값들의 plot을 H의 $(k_1...k_2)$ **하이퍼그램**이라 하고 이를 벡터 함수 $\boldsymbol{h}^{(k1...k2)}$로 표시한다. ***H*의 rank**는 에지들의 cardinality 중 최대값으로 정의된다. H의 rank가 r일 때 (1...r) 하이퍼그램을 ***H*의 하이퍼그램**이라 정의하고 이를 $\boldsymbol{h}$로 표기한다. 위의 하이퍼그램 정의는 하이퍼망의 메모리 구조 또는 코드워드의 분포를 기술하는 한 가지 방법이다. 한편 이 코드의 구조적 특성이 갖는 기능 또는 성능을 평가할 필요가 있으며 이 경우에는 Cardinality 값 k=k_1,k_1+1,...,k_2에 대한 각각의 $H^{(k)}$의 성능지수의 plot

을 H의 **기능적 ($k_1 ... k_2$) 하이퍼그램**이라 하고 이를 벡터 함수 $\boldsymbol{h}_f^{(k_1 ... k_2)}$로 표시한다. H의 기능적 하이퍼그램은 $\boldsymbol{h}_f$ 로 표기한다. 기능적 하이퍼그램과 구별하기 위하여 하이퍼그램을 경우에 따라서는 ``구조적” 하이퍼그램이라고 부를 것이다.

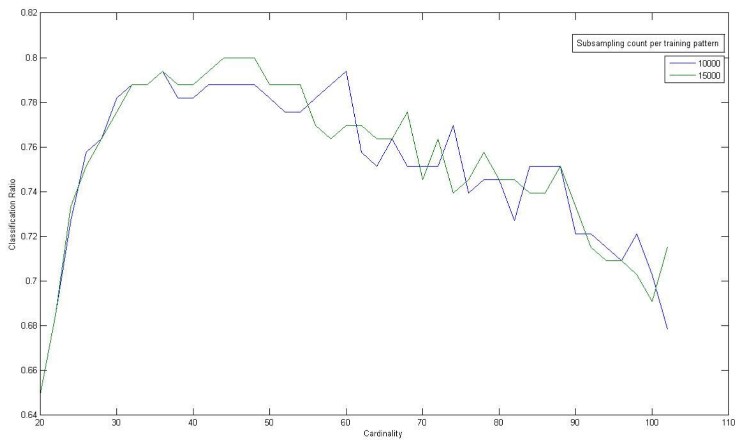


<그림> 숫자 패턴 인식을 위한 하이퍼망에 대한 기능적 하이퍼그램 $\boldsymbol{h}_f$ (하이퍼망의 메모리 용량을 노드수로 고정한 경우). 각각의 실험은 균일한 구조를 갖는 k-하이퍼망을 사용하였으며 k의 값(x 축, 그림에서는 Order로 명시되어 있음)이 커짐에 따라 최대 성능은 향상되는 경향을 보이나 대체로 k=20…30의 구간에서 최대 성능을 보인다.

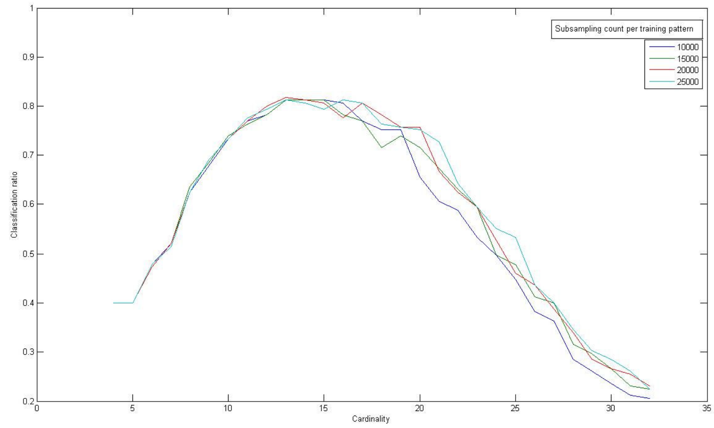


<그림> 숫자 패턴 인식을 위한 하이퍼망에 대한 기능적 하이퍼그램 $\boldsymbol{h}$ (하이퍼망의 메모리 용량을 에지수로 고정한 경우). 각각의 실험은 균일한 구조를 갖는 k-하이퍼망을 사용하였으

며 *k*의 값(*x* 축, 그림에서는 Order로 명시되어 있음)이 커짐에 따라 최대 성능은 향상되는 경향을 보이나 대체로 *k*=25…40의 구간에서 최대 성능을 보인다.



<그림> Face recognition 데이터에 대한 기능적 (20…102) 하이퍼그램. 이미지의 크기는 480 pixels 즉 하이퍼망의 |V| = 480. 하이퍼에지의 크기 *k*=40…50에서 좋은 성능을 보이고 있다.



<그림> Face recognition 데이터에 대한 기능적 (4…32) 하이퍼그램. 이미지의 크기는 64 pixels 즉 하이퍼망의 |V| = 64. 하이퍼에지의 크기 *k*=13…17에서 좋은 성능을 보이고 있다.

4.2.2 Recognition (Classification) Memory Models

위에서 기술한 결합확률분포 모델은 무감독학습의 회상 기억 모델로서 유용하다. 감독학습을 하는 인식 또는 분류 기억장치를 모델링하는 경우 결합확률분포 $P(x, y)$ 대신 조건부 확률분포 $P(y\,|\,x)$를 사용할 수 있다. 즉 변수의 집합 *X* 대신 이제 전체의 변수 집합이 입력변수 *X*와 출력변수 *Y*로 즉 (*X, Y*)로 구성되어 있다고 하면 이에 대한 결합확률분포 모델은

$$P(x, y) = \frac{1}{Z}\exp\left(-\sum_{i=1}^{|E|} w_i \varphi_{E_i}(x, y)\right)$$

이며 이로부터 조건부 확률 분포를 다음과 같이 구할 수 있다.

$$P(y\,|\,x) = \frac{P(x, y)}{P(x)}$$

따라서 확률분포를 모델링하는 하이퍼망 기억구조는 무감독 학습 뿐만 아니라 감독 학습 모델로서 사용될 수 있다.

하이퍼망은 또한 패턴인식 또는 패턴분류와 같은 감독학습 문제를 직접 해결하는 discriminative model로 사용될 수 있다. 이 경우 확률분포를 모델링하는 대신 분류 의사 결정에 필요한 함수값 만을 최대화하도록 학습한다.

$$\begin{aligned} y^* &= \arg\max_y P(y\,|\,x) = \arg\max_\theta f_\theta(x) \\ &= \arg\max_\theta \frac{1}{Z}\exp\left(-\sum_{i=1}^{|E|} w_i \varphi_{E_i}(x)\right) \\ &= \arg\max_\theta \left\{-\sum_{i=1}^{|E|} w_i \varphi_{E_i}(x)\right\} \end{aligned}$$

위 식들에서 Z의 normalization 항이나 $\exp(\cdot)$의 함수 변환은 최대화에 영향을 주지 않기 때문에 간략화되었다. $f_\theta(x)$는 입력 x에 대해서 대응하는 출력 y를 예측하는 함수로서

$$f : \Theta \times X \rightarrow Y$$

$\theta \in \Theta$는 모델의 파라메터 벡터를 나타내며 potential function $\varphi_{E_i}(x)$의 기술에 사용되는 파라메터와 이를 결합하는 가중치 벡터 w를 포함한다. 위의 최대화 식은 다음과 같은 최소화 식과 동일하다.

$$\begin{aligned} y^* &= \arg\min_\theta \left\{ \sum_{i=1}^{|E|} w_i \varphi_{E_i}(x) \right\} \\ &= \arg\min_\theta J_\theta(x) \end{aligned}$$

여기서 $J(x) = \sum_{i=1}^{|E|} w_i \varphi_{E_i}(x)$ 이며 이는 물리적 시스템에서 potential energy에 해당한다.

위의 (3)-(5)에서 기술한 potential function을 사용하여 다양한 형태의 에너지 함수를 정의할 수 있다. 예를 들어, 시그모이드 potential function

$$\varphi_{E_i}(x) = \frac{1}{1 + \exp\left(-\sum_{x_j \in E_i} w_j x_j \right)}$$

을 취할 경우 하이퍼망은 기존의 다층 신경망 모델과 유사한 구조가 된다. 이 하이퍼망의 출력함수는 다음과 같이 주어진다.

$$f(x) = \sum_{i=1}^{|E|} w_i \varphi_{E_i}(x) = \sum_{i=1}^{|E|} \frac{w_i}{1 + \exp\left(-\sum_{x_j \in E_i} w_j x_j \right)}$$

그러나 출력함수는 하이퍼망 모델에서 각 뉴런은 선형적인 가중치의 합 뿐만 아니라 2차 및 그 이상의 고차 가중치 결합을 계산하는 항들을 포함하고 있기 때문에 기존의 신경망을 고차 신경망으로 확장한 것으로 볼 수 있다. 만약 Gaussian potential function을 활성함수로 정의할 경우

$$\varphi_{E_i}(x) = \exp\left(-\frac{\| x^{(i)} - x \|_{E_i}}{\sigma_i} \right)$$

하이퍼망 모델은 radial basis function network과 비슷한 구조를 취한다.

$$f(x) = \sum_{i=1}^{|E|} w_i \varphi_{E_i}(x) = \sum_{i=1}^{|E|} w_i \exp\left(-\frac{\| x^{(i)} - x \|_{E_i}}{\sigma_i} \right)$$

이 구조는 에지에 포함되어 있는 변수들과 매치되는 입력변수들간의 내적을 계산하는 것과 같으며 따라서 입력벡터와 Gaussian basis 함수간의 전체 내적을 계산하는 기존의 RBF를 일반화한 것으로 볼 수 있다. 즉 기존의 RBF 모델은 Gaussian potential function을 취하는 하이퍼망 모델에서 모든 에지의 cardinality가 n인 경우이다.

또 다른 흥미 있는 특수한 경우의 예로서 product potential function을 들 수 있다.

$$\varphi_{E_i}(x) = \prod_{x_j \in E_i} x_j = x_{i_1} x_{i_2} \cdots x_{i_{|E_i|}}$$

이 경우 출력함수는 다음과 같다.

$$f(x) = \sum_{i=1}^{|E|} w_i \varphi_{E_i}(x) = \sum_{i=1}^{|E|} w_i x_{i_1} x_{i_2} \cdots x_{i_{|E_i|}}$$

만약 입출력 변수들이 실수를 취할 경우 $f(x)$는 고차의 다항식을 항으로 가지는 가중치 합이 됨을 주목할 필요가 있다. 즉 이 경우 하이퍼망은 고차 다항식 형태의 추정함수를 표현한다. 만약 입출력 변수들이 이진값을 취하는 경우 $f(x)$는 disjunctions of conjunctions 형태의 논리식을 표현함을 알 수 있다. 그리고 각 disjunct는 가중치가 또한 주어진다. 이러한 weighted DNF 형태(wDNF)의 논리식은 부울 논리식의 brittleness 단점을 개선하여 smoothing한 것이며 이를 기반으로 threshold logic을 구현할 수 있다[Zhang, DNA-2005]. wDNF 식은 또한 기호로 표현된 규칙들의 집합을 표현하는데 적합하며 수치의 결합을 최소화함으로써 학습 결과를 이해하기 쉬운 장점이 있어서 예측 성능 뿐만 아니라 모델의 설명력(explanationability)이나 이해의 용이성(comprehensibility)을 필요로 하는 data mining 응용에 특히 적합하다.