CSC2515 — Assignment #2 Answers

Behdad Esfahbod
993505827

November 9, 2004

1 Adapting Centres in Radial Basis Networks (2%)

- $$\begin{aligned}
\frac{\partial E}{\partial \mathbf{z}_k} &\propto \frac{\partial}{\partial \mathbf{z}_k} \sum_n (y_n - \sum_j w_j h_j(\mathbf{x}_n))^2 \\
&= \sum_n 2(y_n - \sum_j w_j h_j(\mathbf{x}_n)) \frac{\partial}{\partial \mathbf{z}_k} w_k \exp(-\alpha_k (\mathbf{x}_n - \mathbf{z}_k)^\top (\mathbf{x}_n - \mathbf{z}_k)) \\
&= \sum_n 2(y_n - \sum_j w_j h_j(\mathbf{x}_n)) w_k \exp(-\alpha_k (\mathbf{x}_n - \mathbf{z}_k)^\top (\mathbf{x}_n - \mathbf{z}_k)).2\alpha_k (x_n - z_k) \\
&= 4 w_k \alpha_k \sum_n (y_n - \sum_j w_j h_j(\mathbf{x}_n)).h_k(\mathbf{x}_n)(\mathbf{x}_n - \mathbf{z}_k)
\end{aligned}$$

- $$\begin{aligned}
\frac{\partial E}{\partial \log \alpha_k} &\propto \frac{\partial}{\partial \log \alpha_k} \sum_n (y_n - \sum_j w_j h_j(\mathbf{x}_n))^2 \\
&= \sum_n 2(y_n - \sum_j w_j h_j(\mathbf{x}_n))^2 \frac{-\partial}{\partial \log \alpha_k} w_k h_k(\mathbf{x}_n) \\
&= \sum_n 2(y_n - \sum_j w_j h_j(\mathbf{x}_n))^2 \frac{-\partial}{\partial \log \alpha_k} w_k \exp\left(-e^{\log \alpha_k} (\mathbf{x}_n - \mathbf{z}_k)^\top (\mathbf{x}_n - \mathbf{z}_k)\right) \\
&= \sum_n 2(y_n - \sum_j w_j h_j(\mathbf{x}_n))^2 w_k \exp\left(-\alpha_k (\mathbf{x}_n - \mathbf{z}_k)^\top (\mathbf{x}_n - \mathbf{z}_k)\right) .(\mathbf{x}_n - \mathbf{z}_k)^2 \\
&= 2\alpha_k w_k \sum_n (y_n - \sum_j w_j h_j(\mathbf{x}_n)).h_k(\mathbf{x}_n).(\mathbf{x}_n - \mathbf{z}_k)^2
\end{aligned}$$

2 Pseudo-Bayesian Linear Regression (4%)

- We start computing $p(\mathbf{w}|\{\mathbf{x}_n, y_n\}, a, b)$. Since this is a probability distribution, we forget about constants (including the denominator) and finally adjust the solution to sum to one.

$$\begin{aligned} p(\mathbf{w}|\{\mathbf{x}_n, y_n\}, a, b) &= \frac{p(\{\mathbf{x}_n, y_n\}|\mathbf{w})p(\mathbf{w})}{p(\{\mathbf{x}_n, y_n\})} \\ &\propto \prod_n p(\mathbf{x}_n, y_n|\mathbf{w})p(\mathbf{w}) \\ &\propto \exp\left\{\frac{-1}{2b}\sum_n (y_n - \mathbf{w}^\top \mathbf{x}_n)^2 - \frac{1}{2a}\mathbf{w}^2\right\} \\ &\propto \exp\left\{\frac{-1}{2}\left(\mathbf{w}^\top(\frac{\mathbf{I}}{a} + \frac{\sum_n \mathbf{x}_n\mathbf{x}_n^\top}{b})\mathbf{w} - 2\sum_n y_n \mathbf{w}^\top \mathbf{x}_n\right)\right\} \end{aligned}$$

  Now it looks like a gaussian with covariance and mean:

$$\Sigma_{\mathbf{w}} = (\frac{\mathbf{I}}{a} + \frac{\sum_n \mathbf{x}_n\mathbf{x}_n^\top}{b})^{-1}$$

$$\mu_{\mathbf{w}} = \sum_n y_n \mathbf{x}_n \Sigma_{\mathbf{w}}$$

  So:

$$p(\mathbf{w}|\{\mathbf{x}_n, y_n\}, a, b) = \mathcal{N}(\mathbf{w}, \mu_{\mathbf{w}}, \Sigma_{\mathbf{w}}).$$

- We differentiate to find the optimal weights:

$$\frac{\partial}{\partial \mathbf{w}}\cdot = 2\lambda\mathbf{w} - \sum_n 2(y_n - \mathbf{w}^\top \mathbf{x}_n)\mathbf{x}_n$$

3 Regularizing Linear Mixtures of Experts (2%)

- $$\begin{aligned}
\partial \ell^{new}/\partial \mathbf{U}_j &= \sum_n \frac{p(j|\mathbf{x}_n)\partial/\partial \mathbf{U}_j \left[exp(-\frac{1}{2}(\mathbf{y}_n - \mathbf{U}_j\mathbf{x}_n)^\top \Sigma^{-1}(\mathbf{y}_n - \mathbf{U}_j\mathbf{x}_n)\right]}{\sum_j p(j|\mathbf{x}_n)p(\mathbf{y}_n|j,\mathbf{x}_n)} + \partial/\partial \mathbf{U}_j(\lambda \sum_{ijk} U_{ijk}^2) \\
&= \sum_n \frac{p(j|\mathbf{x}_n)p(\mathbf{y}_n|j,\mathbf{x}_n)\Sigma^{-1}(\mathbf{y}_n|\mathbf{U}_j,\mathbf{x}_n)\mathbf{x}_n}{\sum_j p(j|\mathbf{x}_n)p(\mathbf{y}_n|j,\mathbf{x}_n)} + 2\lambda \sum_{ik} U_i jk \frac{\partial}{\partial \mathbf{U}_j}(U_{ijk}) \\
&= \Sigma^{-1} \sum_n p(j|\mathbf{x}_n,\mathbf{y}_n)(\mathbf{y}_n - \mathbf{U}_j\mathbf{x}_n)\mathbf{x}_n + 2\lambda \mathbf{U}_j
\end{aligned}$$

  The last part is deduced by observing that $\frac{\partial}{\partial \mathbf{U}_j} U_{ijk}$ is a matrix of all zeros with a one at entry (i,k).

- To regularize the individual experts, another way is to use ridge regression with each gaussian. Means:

$$\ell = \sum_n \log \sum_j \left[p(j|\mathbf{x}_n)\mathcal{N}(\mathbf{y}_n; \mathbf{U}_j\mathbf{x}_n, \Sigma) + \lambda_j \sum_{ik} U_{ijk}^2 \right]$$

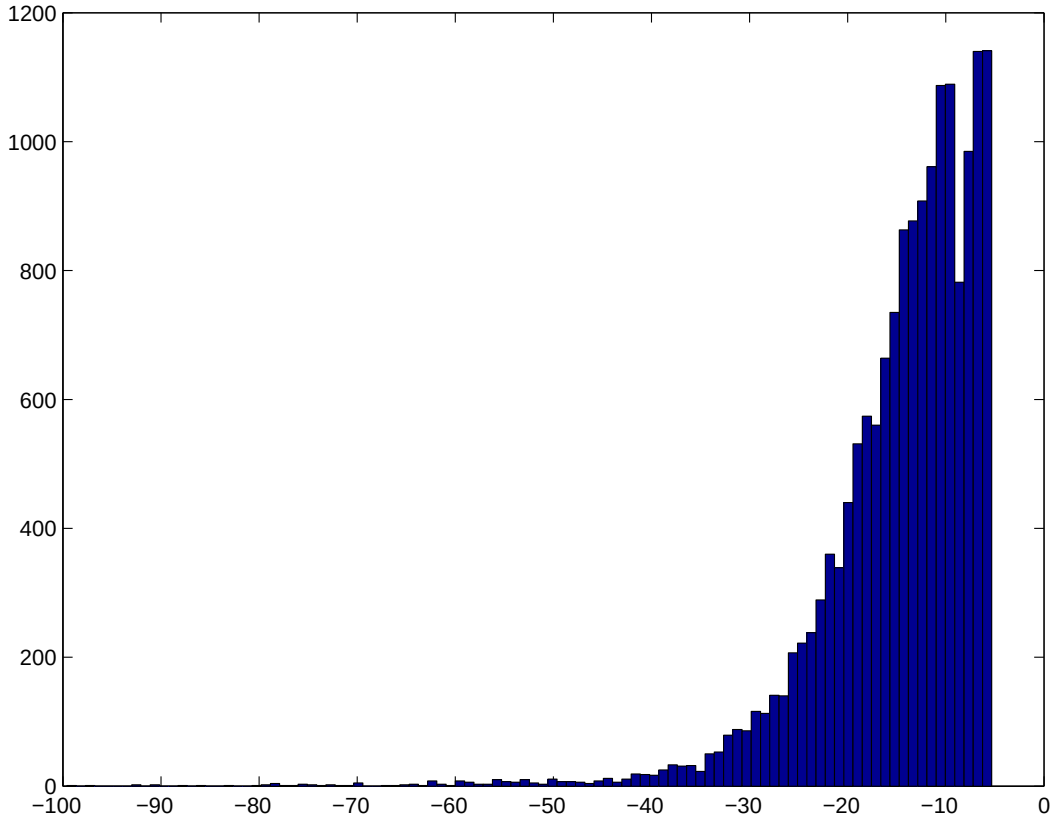
  To regularize the gate, we add $\epsilon \sum v_j^2$ to the cost function:

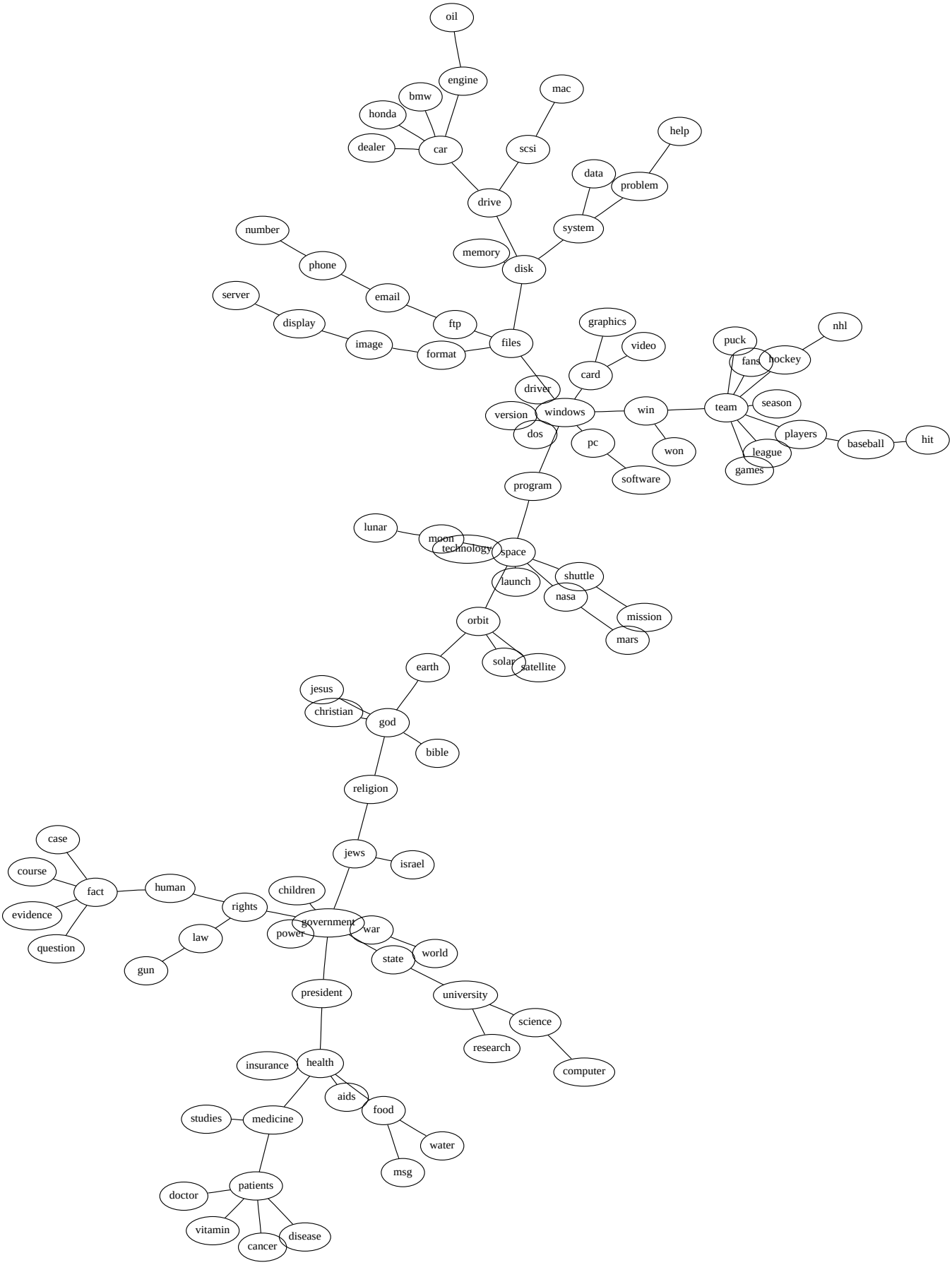
$$\ell = \sum_n \log \sum_j p(j|\mathbf{x}_n)\mathcal{N}(\mathbf{y}_n; \mathbf{U}_j\mathbf{x}_n, \Sigma) + \epsilon \sum_j v_j^2$$

  To regularize the whole architecture, we add some randomly generated pairs of inputs and output (x_i, y_i), considering the range of each variable when generating them.

4 Fully Observed Trees (10%)

Min log likelihood: -99.57562
Max log likelihood: -5.32291
Mean log likelihood: -14.69724
Median log likelihood: -12.89544
Worst log likelihood belongs to document number 10576
This document has words: `car computer course data disk display earth evidence fact files ftp god health help hit human image launch lunar mars mission moon nasa number orbit power program research satellite science solar space studies system technology version war water world`




oil
engine
bmw
mac
honda
help
dealer
car
scsi
data
problem
drive
system
number
memory
phone
disk
server
email
ftp
graphics
nhl
puck
display
image
format
files
video
fans
hockey
card
driver
season
version
windows
win
team
dos
pc
won
players
baseball
hit
league
games
software
program
lunar
moon
technology
space
shuttle
launch
nasa
mission
orbit
mars
solar
satellite
earth
jesus
christian
god
bible
religion
case
jews
israel
course
fact
human
children
evidence
rights
government
war
question
law
power
world
gun
state
president
university
science
research
insurance
health
computer
aids
food
studies
medicine
water
msg
patients
doctor
vitamin
cancer
disease

earth
god
bible
christian
jesus
religion
orbit
satellite
solar
space
launch
moon
lunar
nasa
mars
program
shuttle
mission
technology
government
children
power
president
rights
state
war
jews
israel
health
aids
food
msg
water
insurance
medicine
patients
cancer
disease
doctor
studies
vitamin
fact
case
course
evidence
question
human
law
gun
university
research
science
world
computer
car
bmw
dealer
engine
honda
oil
drive
scsi
mac
disk
memory
data
system
problem
help
graphics
card
video
dos
driver
windows
files
format
ftp
pc
software
version
image
display
server
email
phone
number
win
team
won
fans
games
hockey
nhl
league
players
baseball
hit
puck
season