TAM 598 Lecture 15:

Advanced Topics in Bayesian Linear Regression

Announcements:

- HW 4 covers lectures 13-16; due on Fri Mar 4

updated!

last time: Bayesian Linear Regression

(1) Max Likelihood Estimation (MLE)

(2) Max A Posteriori Estimation (MPE)

(3) Bayesian Linear Regression

Topics

(1) Evidence Approximation to estimate (non-weight) parameters like noise variance and other hyperparameters

(2) Automatic Relevance Determination to select important basis functions

(3) Model quality assessment using standardized errors and quantile-quantile charts

(1) Evidence Approximation

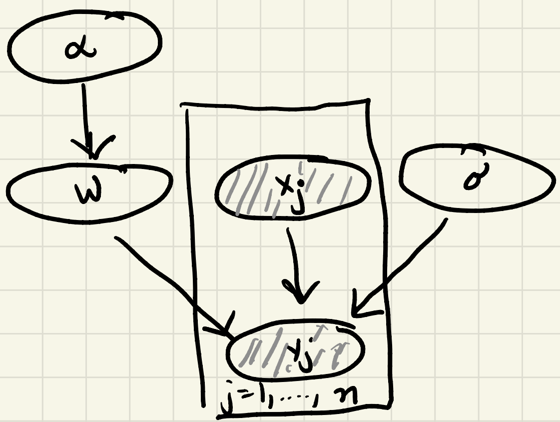
recall: Bayesian linear regression - can separate out epistemic and aleatoric uncertainty, but requires us to choose several parameters by hand:

our model :

data likelihood :

weight prior :

hyper-prior :



Consider the marginal posterior of $\underline{\theta}$:

$$p(\underline{\theta} \mid \underline{x}_{1:n}, \underline{y}_{1:n})$$

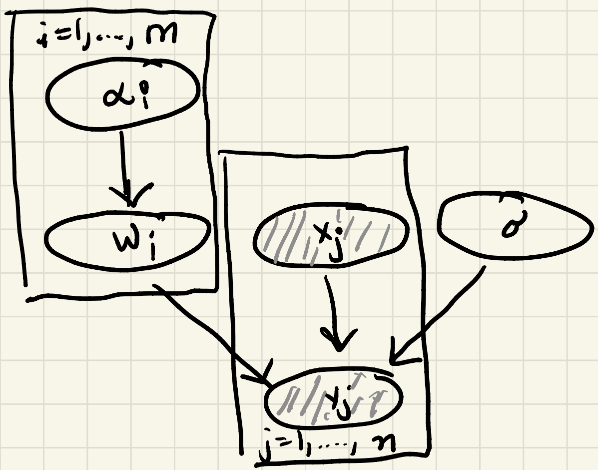
Assume a flat hyperprior $p(\underline{\theta}) \propto 1$.

Use a max a posteriori estimate for $\underline{\theta}$:

(2) Automatic Relevance Determination - how to select which basis functions to keep and which not.

approach: use a different precision α_i for each weight w_i corresponding to each basis function ϕ_i

$$p(w_i \mid \alpha_i) \propto \exp(-\alpha_i w_i^2)$$



in practice:

- evidence approx to find all hyperparameters $\theta = \{\alpha_i, \sigma^2, \ell, \dots\}$
- the precisions α_i of the basis functions that are not needed become very large
- consequently, posteriors of basis function weights collapse to delta function centered at zero

(3) Model Quality Assessment

- need a validation dataset of inputs $\underline{x}^v_{1:n} = (\underline{x}^v_1, \ldots, \underline{x}^v_n)$ and outputs $\underline{y}^v_{1:n} = (y^v_1, y^v_2, \ldots y^v_n)$

statistical diagnostics to compare the predictive distribution to the distribution of the validation dataset

given our gaussian predictive distribution with posterior predictive mean $m(\underline{x})$ and posterior predictive variance $s^2(\underline{x})$, define

standardized errors $$e_i = \frac{y^v_i - m(\underline{x}^v_i)}{s(\underline{x}^v_i)}$$

If our model is correct, these errors should be distributed as $N(0, 1)$ since (if model is correct) $y^v_i \sim N(m(\underline{x}^v_i), \sigma^2(\underline{x}^v_i))$

Approach: calculate $\{e_i\}$, see if $\{e_i\} \sim N(0, 1)$