TAM 598 Lecture 14:

Bayesian Linear Regression

Announcements:

- HW 4 covers lectures 13-16; due on Mar 31 (Mon)
- No class Monday Mar 24th

Today:

(1) Max Likelihood Estimation (MLE)

(2) Max A Posteriori Estimation (MPE)

(3) Bayesian Linear Regression

Least Squares Linear Regression - "traditional"

Can we interpret it from a more "modern" Bayesian or probalistic way?

observations

generalized linear model w/ m basis functions:

model the measurement process using a likelihood function

assume that outcomes of single measurements are Gaussian, with mean $\underline{w}^T \underline{\underline{\phi}}(\underline{x})$
and noise variance σ^2

notation: $N(y \mid \mu, \sigma^2)$ is a pdf

$$p(y) = (2\pi\sigma^2)^{-1/2} \exp\left\{ -\frac{(y-\mu)^2}{2\sigma^2} \right\}$$

for independent measurements, the likelihood of the data factorizes

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

Today:

(1) Max Likelihood Estimation (MLE)

- assume measurement outcomes are gaussian distributed
- find weights $\underline{w}$ and variance σ^2 that maximizes likelihood of measuring the observed data

(2) Max A Posteriori Estimation (MPE)

(3) Bayesian Linear Regression

(I) How do we find the parameters? Maximum likelihood for weights $\underline{w}$, and σ^2

$$\max_{\underline{w},\,\sigma^2} \log p(\underline{y}_{1:n} \mid \underline{x}_{1:n}, \underline{w})$$

$$\max_{\underline{w},\,\sigma^2}$$

(*) So: maximizing likelihood w/r/t $\underline{w}$ is equivalent to minimizing sum of squares! And our weights should therefore satisfy

AND also we can estimate σ^2 as well via max likelihood

Now: we can incorporate uncertainty σ^2 when making predictions

point predictive distribution : $p(y \mid \underline{x}, \underline{w}, \sigma^2) = \mathcal{N}(y \mid \underline{w}^T \phi(\underline{x}), \sigma^2)$

the measured output y is normal distributed around the model (least squares) prediction, with variance σ^2

Today:

(1) Max Likelihood Estimation (MLE)

(2) Max A Posteriori Estimation (MPE)

- assume measurement outcomes are gaussian distributed as before
- assume gaussian <u>prior</u> on weights $p(\underline{w})$ with zero mean, variance α
- assume measurement uncertainty σ^2
- find the weights $\underline{w}$ that maximize the prob. of the Bayesian posterior
- helps to avoid overfitting

(3) Bayesian Linear Regression

II How do we find the parameters? Maximum a posteriori estimates

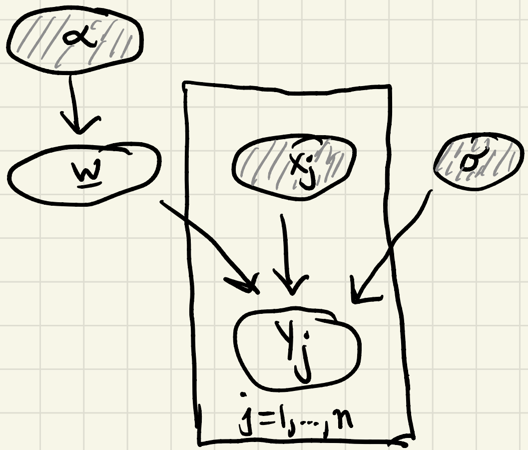
here: maximize the log prob of the posterior rather than the likelihood. Helps avoid overfitting

again: $y(\underline{x}; \underline{w}) =$

new: uncertainty in model parameters described by a prior $\underline{w} \sim p(\underline{w})$

eg) gaussian prior on weights $p(\underline{w} | \alpha) =$

graphically



posterior from Bayes' Rule

$$p(\underline{w} \mid \underline{x}_{1:n}, \underline{y}_{1:n}, \sigma, \alpha) = \frac{p(\underline{y}_{1:n} \mid \underline{x}_{1:n}, \underline{w}, \sigma, \alpha)\, p(\underline{w} \mid \alpha)}{\int p(\underline{y}_{1:n} \mid \underline{x}_{1:n}, \underline{w}', \sigma, \alpha)\, p(\underline{w}' \mid \alpha)\, d\underline{w}'}$$

our state of knowledge about $\underline{w}$, after we see the data

Now assuming σ and α are known.

a point estimate of $\underline{w}$ is

$$\underline{W}_{MPE} =$$

For gaussian likelihood and weight prior :

$$\log p(\underline{w} \mid \underline{x}_{1:n}, \underline{y}_{1:n}, \sigma, \alpha) =$$

maximizing :

Today:

(1) Max Likelihood Estimation (MLE)

(2) Max A Posteriori Estimation (MPE)

(3) Bayesian Linear Regression

- just like (2) above, but now work with the full posterior distribution, not just a point estimate

- allows us to separate epistemic and aleatoric uncertainty

OR another way to see this: our posterior

$$p(\underline{w} \mid \underline{x}_{1:n}, \underline{y}_{1:n}, \sigma^2, \alpha)$$

algebra: rewrite in general form of a gaussian

III How do we find the parameters? Bayesian Linear Regression

similar to II :

again: $y(\underline{x}; \underline{w}) = \underline{w}^T \underline{\phi}(\underline{x})$

$$\underline{y}_{1:n} \mid \underline{x}_{1:n}, \underline{w}, \underline{\sigma}^2 \sim N(\underline{w}^T \underline{\phi}(x), \sigma^2)$$

uncertainty in model parameters described by a prior $\underset{\sim}{w} \sim p(\underline{w})$

but no point estimate. Keep the posterior distribution and work with it directly. Why? Can now quantify the epistemic uncertainty arising from limited # of observations.

posterior predictive distribution: what can we say about y at some new $\underset{\sim}{x}$ after seeing the data?

$$p(y \mid \underset{\sim}{x}, \underset{\sim}{x}_{1:n}, \underset{\sim}{y}_{1:n}, \sigma, \alpha) =$$

for all gaussian priors, this is analytically available

$$p(y \mid \underset{\sim}{x}, \underset{\sim}{x}_{1:n}, \underset{\sim}{y}_{1:n}, \sigma, \alpha) =$$