PROBABILISTIC REGRESSION

ICTP
www.ictp.it

1ST ASSUMPTION: target values are given by a function in terms of the parameters and inputs; and, a noise.

$$t = y(\vec{x},\vec{w}) + \epsilon$$

2ND ASSUMPTION
NOISE IS GAUSSIAN
$\therefore$ LIKELIHOOD IS GAUSSIAN

$$p(t|\vec{x},\vec{w},\beta) \hat{=} \mathcal{N}(\vec{t}\,|\,y(\vec{x},\vec{w}),\beta^{-1})$$

→ Mean of Gaussian is the function
→ Variance of Gaussian is the inverse of precision β.

MATH DEVELOPMENT

CONDITIONAL LIKELIHOOD
1. CONTRIBUTION FROM ALL OBSERVATIONS

$$p(\vec{t}|\vec{x},\vec{w},\beta) = \prod_{n=1}^{N} \mathcal{N}(t_n|y(\vec{x}_n,\vec{w}),\beta^{-1})$$

WITH GENERALIZED LINEAR REGRESSION FUNCTION: $y(\vec{x},\vec{w}) = \vec{w}^T\phi(\vec{x})$

MY NOTATION: $\vec{\phi}_n \equiv \phi(\vec{x}_n)$ $\therefore$

$$p(\vec{t}|\vec{x},\vec{w},\beta) = \prod_{n=1}^{N} \mathcal{N}(t_n|\vec{w}^T\vec{\phi}_n,\beta^{-1})$$

GOAL: MAXIMIZE LIKELIHOOD

$$\underset{\vec{w},\beta}{\mathrm{argmax}}\left[p(\vec{t}|\vec{x},\vec{w},\beta)\right]$$

FOR SIMPLICITY: TAKE LOG:

$$\log(p(\vec{t}|\vec{x},\vec{w},\beta)) = \log\left[\left(\frac{\beta}{2\pi}\right)^{\frac{N}{2}} \cdot \prod_{n=1}^{N} \exp\left(\frac{-\beta}{2}(t_n-\vec{w}^T\vec{\phi}_n)^2\right)\right]$$

$$= \log\left[\left(\frac{\beta}{2\pi}\right)^{\frac{N}{2}}\right] + \sum_{n=1}^{N} \frac{-\beta}{2}(t_n-\vec{w}^T\vec{\phi}_n)^2$$

↳ CONSTANTS, DO NOT PLAY A ROLE IN MAXIMIZING.

$$\underset{\vec{w}}{\mathrm{argmax}}\left[\log(p(\vec{t}|\vec{x},\vec{w},\beta))\right] = \underset{\vec{w}}{\mathrm{argmin}}\left[\sum_{n=1}^{N}\frac{\beta}{2}(t_n-\vec{w}^T\vec{\phi}_n)^2\right]$$

↳ SUM OF SQUARED ERROR.

$$\nabla_{\vec{w}}\left[\sum_{n=1}^{N}\frac{\beta}{2}(t_n-\vec{w}^T\vec{\phi}_n)^2\right] \overset{!}{=} 0 = \sum_{n=1}^{N}\beta(t_n-\vec{w}^T\vec{\phi}_n)\vec{\phi}_n$$

SOLUTION → $\vec{w} = (\phi\phi^T)^{-1}\phi t$

SAME AS LEAST SQUARES REGRESSION!

LEAST-SQUARES REGRESSION IS EQUIVALENT TO MAXIMUM LIKELIHOOD UNDER THE ASSUMPTION OF GAUSSIAN NOISE.

PROBABILISTIC REGRESSION

MAXIMUM - A - POSTERIORI (MAP) ESTIMATION

www.ictp.it

USE RESULTS FROM MAXIMUM LIKELIHOOD WITH GAUSSIAN NOISE; NOW, CONSIDER A GAUSSIAN PRIOR WITH ZERO MEAN AND PRECISION α.

$\vec{x} \Leftrightarrow \phi$

posterior = $\frac{\text{likelihood} \cdot \text{prior}}{\text{normalization (constant)}}$; posterior $\propto$ likelihood $\cdot$ prior

log-likelihood: $\log\left(p(\vec{t}\,|\vec{x},\vec{w},\beta)\right) = \log\left(\frac{\beta}{2\pi}\right)^{\frac{N}{2}} - \sum_{n=1}^{N} \frac{\beta}{2}\left(t_n - \vec{w}^T\vec{x}\right)^2$

prior $:= \sqrt{\frac{\alpha}{2\pi}} \cdot \exp\left(\frac{-1}{2\alpha^{-1}}\|\vec{w}\|_2^2\right)$ $\therefore$ log-prior $= \log\left(\sqrt{\frac{\alpha}{2\pi}}\right) - \frac{\alpha}{2}\|\vec{w}\|_2^2$

so that:

$\log(\text{posterior}) = \text{log-likelihood} + \text{log-prior}$

GOAL: $\underset{\vec{w}}{\operatorname{argmax}} \left[\log(\vec{w}\,|\vec{x},t,\alpha,\beta)\right]$

or minimize with negative $\therefore$

$$\underset{\vec{w}}{\operatorname{argmin}} \left\{ -\left[\frac{N}{2}\log\left(\frac{\beta}{2\pi}\right) + \log\left(\frac{\alpha}{2}\right)\right] + \underbrace{\sum_{n=1}^{N}\frac{\beta}{2}\left(t_n - \vec{w}^T\vec{x}\right)^2 + \frac{\alpha}{2}\|\vec{w}\|_2^2}_{\text{Ridge Regression}} \right\}$$

$$\nabla_w\left[-\log(p(\vec{w}\,|\vec{x},t,\alpha,\beta))\right] = -\sum_{n=1}^{N}\beta\left(t_n - \vec{w}^T\vec{x}\right)\vec{x} + \alpha\vec{w} \overset{!}{=} 0$$

SOLUTION $\rightarrow$ $\vec{w} = \left(\vec{x}\vec{x}^T + \frac{\alpha}{\beta}\mathbb{I}\right)^{-1}\vec{x}\,t_n$

$\hookrightarrow$ Inverse is well-conditioned by the effect of regularization.

RIDGE REGRESSION IS EQUIVALENT TO MAP ESTIMATION WITH GAUSSIAN PRIOR $\rightarrow \mathcal{N}(0, \alpha^{-1})$ AND GAUSSIAN LIKELIHOOD $\rightarrow p(\vec{t}\,|\vec{x},\vec{w},\beta) = \mathcal{N}(t\,|\vec{w}^T\vec{x}, \beta^{-1})$