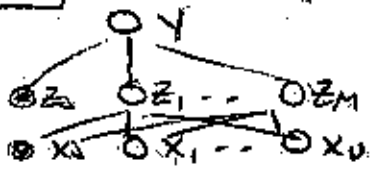
ECE 544 NA
11/6/2013

TODAY

BIAS, VARIANCE, & RESIDUAL



TODAY:

$$\mathcal{E} = \frac{1}{2} \mathop{E}_{\vec{x},t}\left[|y(\vec{x}) - t|^2 \right]$$

$$= \frac{1}{2} E\left[|y(\vec{x}) - \langle t|\vec{x}\rangle|^2 \right] + \frac{1}{2} E\left[|\langle t|\vec{x}\rangle - t|^2 \right]$$

DOES THIS REALLY GO TO ZERO? HOW FAST?

CONDITIONAL VARIANCE OF t GIVEN $\vec{x}$

$$+ E\left[(y(\vec{x}) - \langle t|\vec{x}\rangle)(\langle t|\vec{x}\rangle - t) \right]$$

ASSUME THIS IS ≈ 0

$$y(\vec{x}) = \sum_{j=0}^{M} w_j \phi_j(\vec{x}) \xrightarrow[M\to\infty]{N\to\infty} \langle t|\vec{x}\rangle$$

BUT FOR FIXED N, M, $W = (\Phi^T\Phi)^{-1}\Phi^T T$

$\Phi(n,j) = \phi_j(\vec{x}^{(n)})$ $\quad T(n) = t^{(n)}$

SUPPOSE $E[\phi_i \phi_j] = \delta_{ij}$

THEN $\Phi^T\Phi \approx N I$

$$W \approx \frac{1}{N}\Phi^T T \approx \frac{1}{N}\sum_{n=1}^{N} \vec{\phi}^{(n)} t^{(n)}$$

$$y(\vec{x}) = y(\vec{\phi}) = \vec{\phi}^T\vec{w} = \vec{\phi}^T\left(\frac{1}{N}\sum_{n=1}^{N} t^{(n)}\vec{\phi}^{(n)}\right)$$

WE CAN TALK OF $D = \{\vec{\phi}^{(1)}, t^{(1)}, \ldots, \vec{\phi}^{(N)}, t^{(N)}\}$

IN OTHER WORDS

$$E_D\left[y(\vec{x})\right] - \underset{D}{E}\left[y(\vec{\phi})\right] = \vec{\phi}^T \underbrace{\underset{D}{E}\left[\frac{1}{N}\sum_{n=1}^{N} t^{(n)}\vec{\phi}^{(n)}\right]}_{\text{CORRELATION OF } t,\ \vec{\phi}}$$

IN GENERAL WE CAN TALK OF

$$\underset{D}{E}\left[y(\vec{x})\right] = \int\cdots\int y(\vec{x})\, p(\vec{x}^{(1)}, t^{(1)}, \ldots, \vec{x}^{(N)}, t^{(N)})\, d\vec{x}^{(1)}\cdots dt^{(N)}$$

$$= \int\cdots\int y_1(\vec{x}) \prod_{n=1}^{N} p(\vec{x}^{(n)}, t^{(n)})\, d\vec{x}^{(1)}\cdots dt^{(N)}$$

$$\underset{D,\vec{x}}{E}\left[y(\vec{x})\right] = \int\cdots\int y(\vec{x})\, p(\vec{x}, \underbrace{\vec{x}^{(1)}, t^{(1)}, \ldots, \vec{x}^{(N)}, t^{(N)}}_{\text{TRAINING TOKENS}})\, d\vec{x}\cdots dt^{(N)}$$

(TEST TOKEN)

$$= \int\cdots\int y(\vec{x})\, p(\vec{x}) \prod_{n=1}^{N} p(\vec{x}^{(n)}, t^{(n)})\, d\vec{x}\cdots dt^{(N)}$$

$$\underset{D,\vec{x},t}{E}\left[|y(\vec{x}) - t|^2\right] = \int\cdots\int |y(\vec{x}) - t|^2 \underbrace{p(\vec{x}, t)}_{\text{TEST TOKEN}} \underbrace{\prod_{n=1}^{N} p(\vec{x}^{(n)}, t^{(n)})}_{\text{TRAINING TOKENS}}\, d\vec{x}\cdots dt^{(N)}$$

$$= \underbrace{\underset{D,\vec{x}}{E}\left[\left|y(\vec{x}) - E_D[y(\vec{x})]\right|^2\right]}_{\text{NETWORK VARIANCE}} + \underbrace{\underset{\vec{x}}{E}\left[\left|E_D[y(\vec{x})] - \langle t|\vec{x}\rangle\right|^2\right]}_{\text{NETWORK BIAS}} + \underbrace{\underset{t,\vec{x}}{E}\left[\left|\langle t|\vec{x}\rangle - t\right|^2\right]}_{\text{RESIDUAL (CONDITIONAL) TARGET VARIANCE}}$$

+ CROSS-TERMS THAT GO TO ZERO

VARIANCE

EXAMPLE $y(\vec{x}) = \sum_{j=1}^{M} w_j \phi_j(\vec{x}) = \vec{w}^T \vec{\phi}$

$$\vec{w} = \frac{1}{N} \sum_{n=1}^{N} t^{(n)} \vec{\phi}^{(n)}$$

SUPPOSE $\vec{\phi}^{(n)} \sim \mathcal{N}(0, I)$

AND $P_T(t^{(n)}) = \begin{cases} \frac{1}{2} & t^{(n)} = 1 \\ \frac{1}{2} & t^{(n)} = -1 \end{cases}$

THEN $t^{(n)} \vec{\phi}^{(n)} \sim \mathcal{N}(0, I)$

AND $Var(\vec{w}) = \frac{1}{N^2} \sum_{n=1}^{N} Var(t^{(n)} \vec{\phi}^{(n)}) = \frac{N}{N^2} I$

SO $\boxed{\vec{w} \sim \mathcal{N}(0, \frac{1}{N} I)}$

$$\left(y(\vec{x}) - E_D[y(\vec{x})] \right)$$

$$= \sum_{j=1}^{M} w_j \phi_j(\vec{x}) - E_D\left(\sum_j w_j \phi_j(\vec{x}) \right)$$

$$= \sum_{j=1}^{M} \phi_j(\vec{x}) \underbrace{\left(w_j - E_D[w_j] \right)}_{\mathcal{N}(0, \frac{1}{N})} \sim \mathcal{N}\left(0, \frac{\left(\sum_{j=1}^{M} \phi_j^2(\vec{x})\right)}{N}\right)$$

$$\approx \mathcal{N}(0, \frac{M}{N})$$

SO $E_{D,\vec{x}}\left[\left(y(\vec{x}) - E_D[y(\vec{x})] \right)^2 \right] \approx \frac{M}{N}$

MORE HIDDEN NODES = WORSE VARIANCE

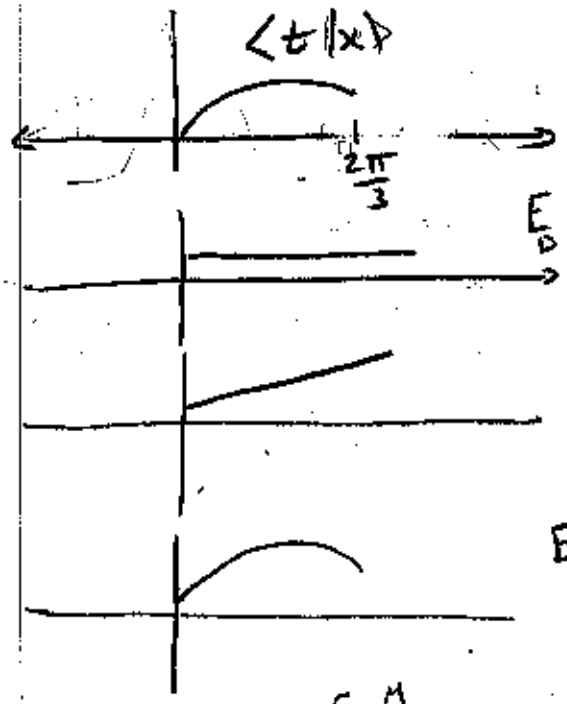
MORE DATA = LOWER VARIANCE

BIAS $\quad \mathop{E}_{x}\left[\left(\mathop{E}_{D}[y(x)] - \langle t|x\rangle\right)^2\right]$

EXAMPLE $\quad x = \text{SCALAR} \sim U(-\Phi, \frac{2\pi}{3})$

$\phi_m(x) = x^m \quad , \quad y(x) = \sum_{m=0}^{M} w_m \phi_m(x)$

$\langle t|x\rangle = \sin(x)$



M=0 $\quad E_D[y(x)] = E_D[w_0]$

M=1 $\quad E_D[y(x)] = E_D[w_0 + w_1 x]$

M=2 $\quad E_D[y(x)] = E_D[w_0 + w_1 x + w_2 x^2]$

$$E_D\left[\sum_{m=0}^{M} w_m x^m\right] = \operatorname{argmin} \mathop{E}_{x}\left[\left(\sin(x) - \sum_{m=0}^{M} w_m x^m\right)^2\right]$$

$\Rightarrow \vec{w} = R^{-1}\vec{r}$

min $\quad \mathop{E}_{x}\left[\left(\sin(x) - \sum_{m=0}^{M} w_m x^m\right)^2\right] = \gamma - \vec{r}^T R^{-1} \vec{r}$

FOR: $\gamma = E[\sin^2(x)]$

$R(m,n) = E[x^{m+n}]$

$\vec{r}(m) = E[x^m \sin(x)]$

MONOTONICALLY NON-INCREASING FUNCTION OF M, INDEPENDENT OF N!

50

$$MSE = \underset{\theta,\vec{x},t}{E}\left[|y(\vec{x}) - t|^2\right]$$

$$= \underset{\vec{x}}{E}\left[BIAS^2(\vec{x})\right] + VARIANCE + RESIDUAL$$