# Pattern Recognition and Machine Learning: Introduction (continued)

Libao Jin

December 8, 2016

# Bayesian Predictive Distribution

$$
\begin{aligned}
p(t|x,\mathbf{x},\mathbf{t}) = \frac{p(t,x,\mathbf{x},\mathbf{t})}{p(x,\mathbf{x},\mathbf{t})} &= \frac{\int p(t,x,\mathbf{x},\mathbf{t},\mathbf{w})d\mathbf{w}}{p(x,\mathbf{x},\mathbf{t})} \\
&= \int \frac{p(t,x,\mathbf{x},\mathbf{t},\mathbf{w})}{p(x,\mathbf{x},\mathbf{t})} d\mathbf{w} \\
&= \int \frac{p(t,x)p(\mathbf{x},\mathbf{t},\mathbf{w})}{p(x)p(\mathbf{x},\mathbf{t})} d\mathbf{w} \\
&= \int \frac{p(t,x)p(\mathbf{w})p(\mathbf{x},\mathbf{t},\mathbf{w})}{p(x)p(\mathbf{w})p(\mathbf{x},\mathbf{t})} d\mathbf{w} \\
&= \int \frac{p(t,x,\mathbf{w})p(\mathbf{x},\mathbf{t},\mathbf{w})}{p(x,\mathbf{w})p(\mathbf{x},\mathbf{t})} d\mathbf{w} \\
&= \int \frac{p(t,x,\mathbf{w})}{p(x,\mathbf{w})} \frac{p(\mathbf{x},\mathbf{t},\mathbf{w})}{p(\mathbf{x},\mathbf{t})} d\mathbf{w} \\
&= \int p(t|x,\mathbf{w})p(\mathbf{w}|\mathbf{x},\mathbf{t})d\mathbf{w}
\end{aligned}
$$

# Bayesian Predictive Distribution

$$p(t|x, \mathbf{x}, \mathbf{t}) = \int p(t|x, \mathbf{w}) p(\mathbf{w}|\mathbf{x}, \mathbf{t}) d\mathbf{w} = \mathcal{N}(t|m(x), s^2(x)),$$

$$\begin{aligned} p(t|x, \mathbf{w}) = p(t|x, \mathbf{w}, \beta) &= \mathcal{N}(t|y(x, \mathbf{w}), \beta^{-1}) \\ &= \frac{1}{(2\pi\beta^{-1})^{1/2}} \exp\left\{ -\frac{1}{2\beta^{-1}} [y(x, \mathbf{w}) - t]^2 \right\} \\ &= \frac{\beta^{1/2}}{(2\pi)^{1/2}} \exp\left\{ -\frac{\beta}{2} [y(x, \mathbf{w}) - t]^2 \right\}, \end{aligned}$$

# Bayesian Predictive Distribution

$$
\begin{aligned}
p(\mathbf{w}|\mathbf{x},\mathbf{t}) =& p(\mathbf{w}|\mathbf{x},\mathbf{t},\alpha,\beta) \propto p(\mathbf{t}|\mathbf{x},\mathbf{w},\beta)p(\mathbf{w}|\alpha) \\
=& \prod_{n=1}^{N} \mathcal{N}(t_n|y(x_n,\mathbf{w}),\beta^{-1}) \cdot \mathcal{N}(\mathbf{w}|\mathbf{0},\alpha^{-1}\mathbf{I}) \\
=& \left[\frac{1}{(2\pi\beta^{-1})^{1/2}}\right]^N \exp\left\{\sum_{n=1}^{N} -\frac{1}{2\beta^{-1}}[y(x_n,\mathbf{w})-t_n]^2\right\} \\
& \cdot \left(\frac{1}{2\pi\alpha^{-1}}\right)^{(M+1)/2} \exp\left\{-\frac{1}{2\alpha^{-1}}\mathbf{w}^T\mathbf{w}\right\} \\
=& \frac{\beta^{N/2}\alpha^{(M+1)/2}}{(2\pi)^{(N+M+1)/2}} \exp\left\{-\frac{\beta}{2}\sum_{n=1}^{N}[y(x_n,\mathbf{w})-t_n]^2 - \frac{\alpha}{2}\mathbf{w}^T\mathbf{w}\right\}
\end{aligned}
$$

# Bayesian Predictive Distribution

$$
\begin{aligned}
& p(t|x, \mathbf{w})p(\mathbf{w}|\mathbf{x}, \mathbf{t}) \\
=& \frac{\beta^{1/2}}{(2\pi)^{1/2}} \exp\left\{-\frac{\beta}{2}[y(x, \mathbf{w}) - t]^2\right\} \\
& \cdot \frac{\beta^{N/2}\alpha^{(M+1)/2}}{(2\pi)^{(N+M+1)/2}} \exp\left\{-\frac{\beta}{2}\sum_{n=1}^{N}[y(x_n, \mathbf{w}) - t_n]^2 - \frac{\alpha}{2}\mathbf{w}^T\mathbf{w}\right\} \\
=& \frac{\beta^{(N+1)/2}\alpha^{(M+1)/2}}{(2\pi)^{(N+M+2)/2}} \\
& \exp\left\{-\frac{\beta}{2}[y(x, \mathbf{w}) - t]^2 - \frac{\beta}{2}\sum_{n=1}^{N}[y(x_n, \mathbf{w}) - t_n]^2 - \frac{\alpha}{2}\mathbf{w}^T\mathbf{w}\right\}
\end{aligned}
$$

# Bayesian Predictive Distribution

$$
\begin{aligned}
&-\frac{\beta}{2}[y(x,\mathbf{w})-t]^2-\frac{\beta}{2}\sum_{n=1}^{N}[y(x_n,\mathbf{w})-t_n]^2-\frac{\alpha}{2}\mathbf{w}^T\mathbf{w}\\
=&-\frac{\beta}{2}\left(\sum_{j=0}^{M}w_jx^j-t\right)^2-\frac{\beta}{2}\sum_{n=1}^{N}\left(\sum_{j=0}^{M}w_jx_n^j-t_n\right)^2-\frac{\alpha}{2}\sum_{j=0}^{M}w_j^2\\
=&-\frac{\beta}{2}\left(\mathbf{w}^T\phi(x)-t\right)^2-\frac{\beta}{2}\sum_{n=1}^{N}\left(\mathbf{w}^T\phi(x_n)-t_n\right)^2-\frac{\alpha}{2}\mathbf{w}^T\mathbf{w}\\
=&-\frac{\beta}{2}\left\{\mathbf{w}^T\left[\phi(x)\phi(x)^T+\sum_{n=1}^{N}\phi(x_n)\phi(x_n)^T+\frac{\alpha}{\beta}I\right]\mathbf{w}\right\}\\
&-\frac{\beta}{2}\left\{-\mathbf{w}^T\left[2\phi(x)t+2\sum_{n=1}^{N}\phi(x_n)t_n\right]+\left(t^2+\sum_{n=1}^{N}t_n^2\right)\right\}
\end{aligned}
$$

# Bayesian Predictive Distribution

$$
\begin{aligned}
p(t|x, \mathbf{x}, \mathbf{t}) &= \int p(t|x, \mathbf{w}) p(\mathbf{w}|\mathbf{x}, \mathbf{t}) d\mathbf{w} \\
&= \frac{1}{[2\pi s^2(x)]^{1/2}} \exp\left\{ \frac{1}{2s^2(x)} [t - m(x)]^2 \right\} \\
&= \mathcal{N}(t|m(x), s^2(x))
\end{aligned}
$$

$$
m(x) = \beta \phi(x)^T \mathbf{S} \sum_{n=1}^{N} \phi(x_n) t_n \qquad s^2(x) = \beta^{-1} + \phi(x)^T \mathbf{S} \phi(x)
$$

$$
\mathbf{S}^{-1} = \alpha \mathbf{I} + \beta \sum_{n=1}^{N} \phi(x_n) \phi(x_n)^T \qquad \phi(x_n) = (x_n^0, x_n^1, \ldots, x_n^M)^T
$$

# Bayesian Predictive Distribution

$$p(t|x, \mathbf{x}, \mathbf{t}) = \int p(t|x, \mathbf{w}) p(\mathbf{w}|\mathbf{x}, \mathbf{t}) d\mathbf{w} = \mathcal{N}(t|m(x), s^2(x))$$