- Sum-of-Squares Error Function:

$$\boxed{E(\mathbf{w}) = \frac{1}{2}\sum_{n=1}^{N}\{y(x_n,\mathbf{w}) - t_n\}^2}$$

- Root-Mean-Square (RMS) Error:

$$E_{\mathrm{RMS}} = \sqrt{2E(\mathbf{w}^\star)/N}$$

- Regularization: Penalize large coefficient values

$$\boxed{\widetilde{E}(\mathbf{w}) = \frac{1}{2}\sum_{n=1}^{N}\{y(x_n,\mathbf{w}) - t_n\}^2 + \frac{\lambda}{2}\|\mathbf{w}\|^2}$$

- Bayes' Theorem:

$$\boxed{p(Y|X) = \frac{p(X|Y)p(Y)}{p(X)}} \qquad \boxed{p(X) = \sum_Y p(X|Y)p(Y)}$$

- Expectations:

$$\boxed{\mathbb{E}[f] = \sum_x p(x)f(x)}$$

- Conditional Expectation (discrete)

$$\boxed{\mathbb{E}_x[f|y] = \sum_x p(x|y)f(x)}$$

- Approximate Expectation (discrete and continuous)

$$\boxed{\mathbb{E}[f] \simeq \frac{1}{N}\sum_{n=1}^{N} f(x_n)}$$

- Variance:

$$\boxed{\mathrm{var}[f] = \mathbb{E}\left[\left(f(x) - \mathbb{E}[f(x)]\right)^2\right] = \mathbb{E}[f(x)^2] - \mathbb{E}[f(x)]^2}$$

- Covariances:

$$\boxed{\begin{aligned}
\mathrm{cov}[x,y] &= \mathbb{E}_{x,y}\left[\{x - \mathbb{E}[x]\}\{y - \mathbb{E}[y]\}\right] \\
&= \mathbb{E}_{x,y}[xy] - \mathbb{E}[x]\mathbb{E}[y] \\
\\
\mathrm{cov}[\mathbf{x},\mathbf{y}] &= \mathbb{E}_{\mathbf{x},\mathbf{y}}\left[\{\mathbf{x} - \mathbb{E}[\mathbf{x}]\}\{\mathbf{y}^{\mathrm{T}} - \mathbb{E}[\mathbf{y}^{\mathrm{T}}]\}\right] \\
&= \mathbb{E}_{\mathbf{x},\mathbf{y}}[\mathbf{x}\mathbf{y}^{\mathrm{T}}] - \mathbb{E}[\mathbf{x}]\mathbb{E}[\mathbf{y}^{\mathrm{T}}]
\end{aligned}}$$

- The Gaussian Distribution:

$$\mathcal{N}\left(x|\mu,\sigma^2\right) = \frac{1}{(2\pi\sigma^2)^{1/2}}\exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\}$$

- Gaussian Mean and Variance:

$$\mathbb{E}[x] = \int_{-\infty}^{\infty}\mathcal{N}\left(x|\mu,\sigma^2\right)x\,\mathrm{d}x = \mu$$

$$\mathbb{E}[x^2] = \int_{-\infty}^{\infty}\mathcal{N}\left(x|\mu,\sigma^2\right)x^2\,\mathrm{d}x = \mu^2+\sigma^2$$

$$\mathrm{var}[x] = \mathbb{E}[x^2]-\mathbb{E}[x]^2 = \sigma^2$$

- The Multivariate Gaussian:

$$\mathcal{N}(\mathbf{x}|\boldsymbol{\mu},\boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{D/2}}\frac{1}{|\boldsymbol{\Sigma}|^{1/2}}\exp\left\{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right\}$$

- Maximum (Log) Likelihood:

$$\ln p\left(\mathbf{x}|\mu,\sigma^2\right) = -\frac{1}{2\sigma^2}\sum_{n=1}^{N}(x_n-\mu)^2-\frac{N}{2}\ln\sigma^2-\frac{N}{2}\ln(2\pi)$$

$$\mu_{\mathrm{ML}} = \frac{1}{N}\sum_{n=1}^{N}x_n \qquad \sigma^2_{\mathrm{ML}} = \frac{1}{N}\sum_{n=1}^{N}(x_n-\mu_{\mathrm{ML}})^2$$

- Entropy:

$$\mathrm{H}[x] = -\sum_{x}p(x)\log_2 p(x)$$

- Conditional Entropy:

$$\mathrm{H}[\mathbf{y}|\mathbf{x}] = -\iint p(\mathbf{y},\mathbf{x})\ln p(\mathbf{y}|\mathbf{x})\,\mathrm{d}\mathbf{y}\,\mathrm{d}\mathbf{x}$$

- Joint Entropy:

$$\mathrm{H}[\mathbf{x},\mathbf{y}] = \mathrm{H}[\mathbf{y}|\mathbf{x}]+\mathrm{H}[\mathbf{x}]$$

- Mutual information:

$$\mathrm{I}[\mathbf{x},\mathbf{y}] = \mathrm{H}[\mathbf{x}]-\mathrm{H}[\mathbf{x}|\mathbf{y}] = \mathrm{H}[\mathbf{y}]-\mathrm{H}[\mathbf{y}|\mathbf{x}]$$

- Binomial Distribution:

$$\mathrm{Bin}(m|N,\mu) = \binom{N}{m}\mu^m(1-\mu)^{N-m}$$

$$\mathbb{E}[m] \equiv \sum_{m=0}^{N} m\mathrm{Bin}(m|N,\mu) = N\mu$$

$$\mathrm{var}[m] \equiv \sum_{m=0}^{N} (m-\mathbb{E}[m])^2\,\mathrm{Bin}(m|N,\mu) = N\mu(1-\mu)$$

- Beta Distribution:

$$\begin{aligned}
\mathrm{Beta}(\mu|a,b) &= \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\mu^{a-1}(1-\mu)^{b-1} \\
\mathbb{E}[\mu] &= \frac{a}{a+b} \\
\mathrm{var}[\mu] &= \frac{ab}{(a+b)^2(a+b+1)}
\end{aligned}$$

- The Multinomial Distribution:

$$\begin{aligned}
\mathrm{Mult}(m_1, m_2, \ldots, m_K|\boldsymbol{\mu}, N) &= \binom{N}{m_1 m_2 \ldots m_K}\prod_{k=1}^{K}\mu_k^{m_k} \\
\mathbb{E}[m_k] &= N\mu_k \\
\mathrm{var}[m_k] &= N\mu_k(1-\mu_k) \\
\mathrm{cov}[m_j m_k] &= -N\mu_j\mu_k
\end{aligned}$$

- The Dirichlet Distribution:

$$\mathrm{Dir}(\boldsymbol{\mu}|\boldsymbol{\alpha}) = \frac{\Gamma(\alpha_0)}{\Gamma(\alpha_1)\cdots\Gamma(\alpha_K)}\prod_{k=1}^{K}\mu_k^{\alpha_k-1}$$

$$\alpha_0 = \sum_{k=1}^{K}\alpha_k$$

$$p(Y|X) = \frac{p(X|Y)p(Y)}{p(X)}$$