1.

$$y(x, \mathbf{w}) \ = \ w_0 + w_1 x + w_2 x^2 + \cdots + w_M x^M \ = \sum_{j=0}^{M} w_j \big(x_n\big)^j$$

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

$$\Rightarrow E(\mathbf{w}) = \frac{1}{2} \sum_{n=1}^{N} \Big[\sum_{j=0}^{M} w_j \big(x_n\big)^j - t_n \Big]^2$$

$$\frac{\partial E}{\partial w_i} = \frac{1}{2} \sum_{n=1}^{N} 2 \Big[\sum_{j=0}^{M} w_j (x_n)^j - t_n \Big] \frac{\partial}{\partial w_i} \Big[\sum_{j=0}^{M} w_j (x_n)^j - t_n \Big]$$

$$= \sum_{n=1}^{N} \Big[\sum_{j=0}^{M} w_j (x_n)^j - t_n \Big] \sum_{j=0}^{M} (x_n)^j \, \frac{\partial w_i}{\partial w_j} \left(\frac{\partial w_i}{\partial w_j} = \delta_{ij} = \begin{cases} 1, & \text{if } j = i, \\ 0, & \text{if } j \neq i \end{cases} \right)$$

$$= \sum_{n=1}^{N} \Big[\sum_{j=0}^{M} w_j (x_n)^j - t_n \Big] (x_n)^i$$

$$\frac{\partial E}{\partial w_i} = 0$$

$$\Rightarrow \sum_{n=1}^{N} \Big[\sum_{j=0}^{M} w_j (x_n)^j - t_n \Big] (x_n)^i = 0$$

$$\Rightarrow \sum_{n=1}^{N} \sum_{j=0}^{M} w_j (x_n)^j (x_n)^i - \sum_{n=1}^{N} t_n (x_n)^i = 0$$

$$\Rightarrow \sum_{n=1}^{N} \sum_{j=0}^{M} w_j (x_n)^{j+i} - \sum_{n=1}^{N} t_n (x_n)^i = 0$$

$$\Rightarrow \sum_{j=0}^{M} w_j \sum_{n=1}^{N} (x_n)^{j+i} - \sum_{n=1}^{N} t_n (x_n)^i = 0$$

$$\Rightarrow \sum_{j=0}^{M} w_j \sum_{n=1}^{N} (x_n)^{j+i} = \sum_{n=1}^{N} t_n (x_n)^i$$

$$\Rightarrow \sum_{j=0}^{M} A_{ij} w_j = T_i \left(A_{ij} = \sum_{n=1}^{N} (x_n)^{j+i}, \quad T_i = \sum_{n=1}^{N} t_n (x_n)^i \right)$$

In matrix notation:

$$\mathbf{A}\,\mathbf{w} = \mathbf{T}$$

$$\Rightarrow \mathbf{w} = \mathbf{A}^{-1}\mathbf{T}$$

2.

$$y(x, \mathbf{w}) \;=\; w_0 + w_1 x + w_2 x^2 + \cdots + w_M x^M \;= \sum_{j=0}^{M} w_j \big(x_n\big)^j$$

$$\begin{aligned}
\tilde{E}(\mathbf{w}) &= \frac{1}{2}\sum_{n=1}^{N}\Big[y(x_n, \mathbf{w}) - t_n\Big]^2 + \frac{\lambda}{2}\|\mathbf{w}\|^2 \\
\Rightarrow \tilde{E}(\mathbf{w}) &= \frac{1}{2}\sum_{n=1}^{N}\Big[\sum_{j=0}^{M} w_j \big(x_n\big)^j - t_n\Big]^2 + \frac{\lambda}{2}\sum_{j=0}^{M} w_j^2
\end{aligned}$$

$$\begin{aligned}
\frac{\partial \tilde{E}}{\partial w_i} &= \frac{\partial}{\partial w_i}\frac{1}{2}\sum_{n=1}^{N}\Big[\sum_{j=0}^{M} w_j \big(x_n\big)^j - t_n\Big]^2 + \frac{\partial}{\partial w_i}\frac{\lambda}{2}\sum_{j=0}^{M} w_j^2 \\
&= \frac{1}{2}\sum_{n=1}^{N} 2\Big[\sum_{j=0}^{M} w_j \big(x_n\big)^j - t_n\Big]\frac{\partial}{\partial w_i}\Big[\sum_{j=0}^{M} w_j \big(x_n\big)^j - t_n\Big] + \frac{\lambda}{2}\frac{\partial}{\partial w_i}\sum_{j=0}^{M} w_j^2 \\
&= \frac{1}{2}\sum_{n=1}^{N} 2\Big[\sum_{j=0}^{M} w_j \big(x_n\big)^j - t_n\Big]\sum_{j=0}^{M}(x_n)^j\,\frac{\partial w_i}{\partial w_j} + \frac{\lambda}{2}\frac{\partial}{\partial w_i}\sum_{j=0}^{M} w_j^2\,\left(\frac{\partial w_i}{\partial w_j} = \delta_{ij} = (x_n)^i \text{ if } j = i\right) \\
&= \sum_{n=1}^{N}\Big[\sum_{j=0}^{M} w_j \big(x_n\big)^j - t_n\Big](x_n)^i + \lambda w_i
\end{aligned}$$

$$\begin{aligned}
&\frac{\partial \tilde{E}}{\partial w_i} = 0 \\
&\Rightarrow \sum_{n=1}^{N}\Big[\sum_{j=0}^{M} w_j \big(x_n\big)^j - t_n\Big](x_n)^i + \lambda w_i = 0 \\
&\Rightarrow \sum_{n=1}^{N}\sum_{j=0}^{M} w_j \big(x_n\big)^j (x_n)^i - \sum_{n=1}^{N}(x_n)^i t_n + \lambda w_i = 0 \\
&\Rightarrow \sum_{n=1}^{N}\sum_{j=0}^{M} w_j \big(x_n\big)^{j+i} - \sum_{n=1}^{N}(x_n)^i t_n + \lambda w_i = 0 \\
&\Rightarrow \sum_{j=0}^{M} w_j \sum_{n=1}^{N}\big(x_n\big)^{j+i} - \sum_{n=1}^{N}(x_n)^i t_n + \lambda w_i = 0 \\
&\Rightarrow \sum_{j=0}^{M} w_j \sum_{n=1}^{N}\big(x_n\big)^{j+i} + \lambda w_i = \sum_{n=1}^{N}(x_n)^i t_n \\
&\Rightarrow \sum_{j=0}^{M} A_{ij} w_j + \lambda w_i = T_i \left(A_{ij} = \sum_{n=1}^{N}(x_n)^{j+i}, \quad T_i = \sum_{n=1}^{N} t_n (x_n)^i\right)
\end{aligned}$$

In matrix notation:

$$\begin{aligned}
&\mathbf{A}\,\mathbf{w} + \lambda\,\mathbf{w} = \mathbf{T} \\
&\Rightarrow (\mathbf{A} + \lambda\,\mathbf{I})\,\mathbf{w} = \mathbf{T} \\
&\Rightarrow \mathbf{w} = (\mathbf{A} + \lambda\,\mathbf{I})^{-1}\,\mathbf{T}
\end{aligned}$$

4.

$$E[L(\mathbf{t}, \mathbf{y}(x))] = \int\int \|\mathbf{y}(x) - \mathbf{t}\|^2 p(x,t)\,dx\,dt.$$

Often, this is simply written as: $\quad E[L] = \int\int \|\mathbf{y}(x) - \mathbf{t}\|^2 p(x,t)\,dx\,dt.$

Since $\|\mathbf{y}(x) - \mathbf{t}\|^2 = (\mathbf{y}(x) - \mathbf{t})^\top(\mathbf{y}(x) - \mathbf{t}),$

we can expand it as: $\quad \|\mathbf{y}(x) - \mathbf{t}\|^2 = \sum_{i=1}^{M} \big[\mathbf{y}_i(x) - \mathbf{t}_i\big]^2.$

Hence: $\quad E[L] = \int\int \sum_{i=1}^{M} \big[\mathbf{y}_i(x) - \mathbf{t}_i\big]^2 p(x,t)\,dx\,dt.$

$$\begin{aligned}\delta E[L] &= \frac{\delta}{\delta \mathbf{y}_j(x)} \int\int \sum_{i=1}^{M} \big[\mathbf{y}_i(x) - \mathbf{t}_i\big]^2 p(x,t)\,dx\,dt \\ &= \int\int \frac{\delta}{\delta \mathbf{y}_j(x)} \left[\sum_{i=1}^{M} \big(\mathbf{y}_i(x) - \mathbf{t}_i\big)^2\right] p(x,t)\,dx\,dt.\end{aligned}$$

Inside the summation, the only term that depends on $\mathbf{y}_j(x)$ is $\big(\mathbf{y}_j(x) - \mathbf{t}_j\big)^2$. Its derivative is:

$$\frac{\delta}{\delta \mathbf{y}_j(x)} \left[\big(\mathbf{y}_j(x) - \mathbf{t}_j\big)^2\right] = 2\big[\mathbf{y}_j(x) - \mathbf{t}_j\big].$$

Thus:

$$\frac{\delta}{\delta \mathbf{y}_j(x)} \left[\sum_{i=1}^{M} \big(\mathbf{y}_i(x) - \mathbf{t}_i\big)^2\right] = 2\big[\mathbf{y}_j(x) - \mathbf{t}_j\big].$$

Substituting this back, we have:

$$\frac{\delta E[L]}{\delta \mathbf{y}_j(x)} = \int\int 2\big[\mathbf{y}_j(x) - \mathbf{t}_j\big] p(x,t)\,dx\,dt.$$

$$\frac{\delta E[L]}{\delta \mathbf{y}_j(x)} = 2\int \big[\mathbf{y}_j(x) - \mathbf{t}_j\big] p(t \mid x)\,dt,$$

where $p(t \mid x) = \frac{p(x,t)}{p(x)}$.

$$\frac{\delta E[L]}{\delta \mathbf{y}_j(x)} = 0 \implies \int 2\big[\mathbf{y}_j(x) - \mathbf{t}_j\big] p(x,t)\,dt = 0.$$

Divide out the constant 2:

$$\int \big[\mathbf{y}_j(x) - \mathbf{t}_j\big] p(x,t)\,dt = 0.$$

Distribute inside the integral:

$$\mathbf{y}_j(x) \int p(x,t)\,dt - \int \mathbf{t}_j p(x,t)\,dt = 0.$$

Since $\int p(x,t)\,dt = p(x)$, we get:

$$\mathbf{y}_j(x)p(x) = \int \mathbf{t}_j p(x,t)\,dt.$$

$$\mathbf{y}(x) = \frac{\int \mathbf{t}\,p(t,x)\,dt}{\int p(t,x)\,dt} = \int \mathbf{t}\,p(t \mid x)\,dt.$$

$$\mathbf{y}_j(x) = \frac{\int \mathbf{t}_j\,p(x,t)\,dt}{\int p(x,t)\,dt} = \frac{\int \mathbf{t}_j\,p(x,t)\,dt}{p(x)} = \mathbb{E}[\mathbf{t}_j \mid x].$$

5.

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

$$\binom{N}{m} + \binom{N}{m-1} = \binom{N+1}{m}$$

$$\binom{N}{m} = \frac{N!}{m!\,(N-m)!}$$

$$\binom{N}{m-1} = \frac{N!}{(m-1)!\,[N-(m-1)]!} = \frac{N!}{(m-1)!\,(N-m+1)!}$$

$$\begin{aligned}
\binom{N}{m} + \binom{N}{m-1} &= \frac{N!}{m!(N-m)!} + \frac{N!}{(m-1)!(N-m+1)!} \\
&= \frac{N!}{m\,(m-1)!(N-m)!} + \frac{N!}{(m-1)!\,(N-m+1)\,(N-m)!} \\
&\quad \text{(As } m! = m\,(m-1)! \text{ and } (N-m+1)! = (N-m+1)\,(N-m)!\text{)} \\
&= \frac{N!}{(N-m)!(m-1)!}\,\frac{1}{m} + \frac{N!}{(N-m)!(m-1)!}\,\frac{1}{N-m+1} \\
&= \frac{N!}{(N-m)!(m-1)!}\left(\frac{1}{m} + \frac{1}{N-m+1}\right) \\
&= \frac{N!}{(N-m)!(m-1)!}\left(\frac{N-m+1+m}{m(N-m+1)}\right) \\
&= \frac{N!}{(N-m)!(m-1)!}\left(\frac{N+1}{m(N-m+1)}\right).
\end{aligned}$$

$$\begin{aligned}
\binom{N+1}{m} &= \frac{(N+1)!}{m!\,\big[(N+1)-m\big]!} = \frac{(N+1)!}{m!\,(N+1-m)!} \\
&= \frac{(N+1)\,N!}{m\,(m-1)!\,(N+1-m)\,(N-m)!} \\
&= \frac{N!}{(N-m)!\,(m-1)!}\,\frac{N+1}{m\,(N+1-m)}.
\end{aligned}$$

Since the expressions for $\binom{N}{m} + \binom{N}{m-1}$ and $\binom{N+1}{m}$ match, identity is proven:

$$\binom{N}{m} + \binom{N}{m-1} \; = \; \binom{N+1}{m}.$$

$$(1+x)^N = \sum_{m=0}^{N} \binom{N}{m} x^m, \quad \text{where } \binom{N}{m} = \frac{N!}{m!(N-m)!}.$$

Base Case: N = 0
When $N = 0$, the left-hand side is

$$(1+x)^0 = 1.$$

The right-hand side is

$$\sum_{m=0}^{0} \binom{0}{m} x^m = \binom{0}{0} x^0 = 1,$$

because $\binom{0}{0} = 1$ and $x^0 = 1$.
Inductive Step: From N to $N+1$
Start with the left-hand side at $N+1$:

$$(1+x)^{N+1} = (1+x)(1+x)^N.$$

By the inductive hypothesis, we know that $(1+x)^N$ is

$$\sum_{m=0}^{N} \binom{N}{m} x^m.$$

Substitute this into the equation:

$$(1+x)^{N+1} = (1+x) \sum_{m=0}^{N} \binom{N}{m} x^m.$$

$$(1+x)^{N+1} = \sum_{m=0}^{N} \binom{N}{m} x^m + \sum_{m=0}^{N} \binom{N}{m} x^{m+1}$$

$$(1+x)^{N+1} = \binom{N}{0} x^0 + \sum_{m=1}^{N} \binom{N}{m} x^m + \sum_{m=1}^{N+1} \binom{N}{m-1} x^m$$

At $m = 0$, we only have

$$\binom{N}{0} x^0 = 1.$$

At $m = N+1$, we only have

$$\binom{N}{(N+1)-1} x^{N+1} = \binom{N}{N} x^{N+1} = x^{N+1}.$$

$$(1+x)^{N+1} = \binom{N}{0} x^0 + \sum_{m=1}^{N} \left[\binom{N}{m} + \binom{N}{m-1} \right] x^m + \binom{N}{N} x^{N+1} \tag{1}$$

Use this to simplify each bracketed pair:

$$\binom{N}{m} + \binom{N}{m-1} = \binom{N+1}{m}.$$

Thus, for $1 \leq m \leq N$, each pair becomes

$$\binom{N+1}{m} x^m.$$

Also note:

$$\binom{N}{0} = 1, \quad \binom{N+1}{0} = 1, \quad \binom{N}{N} = 1, \quad \binom{N+1}{N+1} = 1.$$

Hence, the expression reorganizes as:

$$(1+x)^{N+1} = \binom{N+1}{0} x^0 + \sum_{m=1}^{N} \binom{N+1}{m} x^m + \binom{N+1}{N+1} x^{N+1}.$$

But this is simply the sum from $m = 0$ to $m = N + 1$:

$$(1+x)^{N+1} = \sum_{m=0}^{N+1} \binom{N+1}{m} x^m.$$

Base Case: $N = 0$: Verified

$$(1+x)^0 = 1.$$

Inductive Step: Shown that if

$$(1+x)^N = \sum_{m=0}^{N} \binom{N}{m} x^m,$$

then

$$(1+x)^{N+1} = \sum_{m=0}^{N+1} \binom{N+1}{m} x^m.$$

By the Principle of Mathematical Induction, we have proved that, for all integers $N \geq 0$,

$$(1+x)^N = \sum_{m=0}^{N} \binom{N}{m} x^m.$$

Binomial Distribution:

$$\mathrm{Bin}(m \mid N, \mu) = \binom{N}{m} \mu^m (1-\mu)^{N-m}, \quad m = 0, 1, \ldots, N.$$

We want to show that this distribution sums to 1:

$$\sum_{m=0}^{N} \binom{N}{m} \mu^m (1-\mu)^{N-m} = 1.$$

2. Binomial Theorem:
From the binomial theorem, we know:

$$(1+x)^N = \sum_{m=0}^{N} \binom{N}{m} x^m \quad \text{for any real (or complex) } x.$$

3. Applying the Theorem to the Binomial Distribution:
Choose $x = \frac{\mu}{1-\mu}$. However, an even simpler approach is to directly observe that:

$$\mu + (1-\mu) = 1.$$

Hence, if we set $\alpha = \mu$ and $\beta = 1 - \mu$ in the binomial theorem, we get:

$$(\alpha + \beta)^N = \sum_{m=0}^{N} \binom{N}{m} \alpha^m \beta^{N-m}.$$

Then $\alpha + \beta = \mu + (1 - \mu) = 1$, so the left-hand side is:

$$1^N = 1.$$

Meanwhile, the right-hand side is:

$$\sum_{m=0}^{N} \binom{N}{m} \mu^m (1-\mu)^{N-m}.$$

Putting it together:

$$1 = \sum_{m=0}^{N} \binom{N}{m} \mu^m (1-\mu)^{N-m}.$$

4. Conclusion:
Thus, we have shown that:

$$\sum_{m=0}^{N} \mathrm{Bin}(m \mid N, \mu) = \sum_{m=0}^{N} \binom{N}{m} \mu^m (1-\mu)^{N-m} = 1,$$

demonstrating that the binomial distribution is properly normalized.