fitting polynomials

$y = w_0 + w_1 x + w_2 x^2 + w_3 x^3 + \cdots + w_9 x^9$

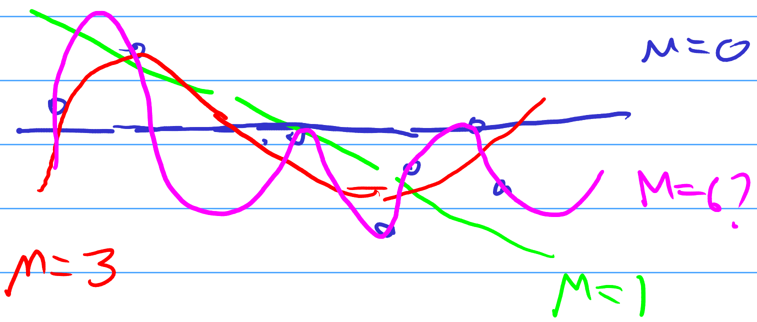
$(x_1, t_1), (x_2, t_2), \ldots, (x_N, t_N)$

1 2 9 ⋯ 9n

$$w_0^{(k+1)} \Leftarrow w_0^k - \frac{\partial E}{\partial w_0}$$

We need to determine the correct polynomial

0th order polynomial



RMSE

Root mean square error

Although M=6 fits the best for this training data, i.e. error → 0

M=3 is actually the better fit, since it may generalize well

@ M=6

Training error may go down, but unseen or testing error is very high

Such models are overfitting models: fits training data very well but cant fit testing data

for larger degrees when you observe the coeffecients, they get quite large and change sign.

Regularized

$$\frac{\partial E}{\partial w_0} = (x_1, t_1)\ (x_2, t_2)\ \cdots\ (x_N, t_n)$$

$$y = w_0 + w_1 x + w_2 x^2 + \cdots + w_9 x^9$$

$$E(w) = \frac{1}{2N} \sum_{n=1}^{N} \left((w_0 + w_1 x_n + w_2 x_n^2 + \cdots w_9 x_n^9) - t_n \right)^2$$

$$\min(E(w))$$

$$\text{s.t. } w_0^2 + w_1^2 + \cdots + w_9^2 = 0.8$$

so we dont get crazy large coefficients this way

$w_0 = 1000$

$w_1 = -3000$

$$\min E(w)$$
$$\sum_{i=0}^{9} w_i^2 \leq \eta$$

$$\boxed{\begin{array}{l} \min E(w) \\ \|w\| \leq E \end{array}}$$

$$\|w\| = \sqrt{w_0^2 + w_1^2 \cdots w_9^2}$$

So the freedom of ws has been curtailed

$\Rightarrow$ ω_s have been <u>regularized</u>

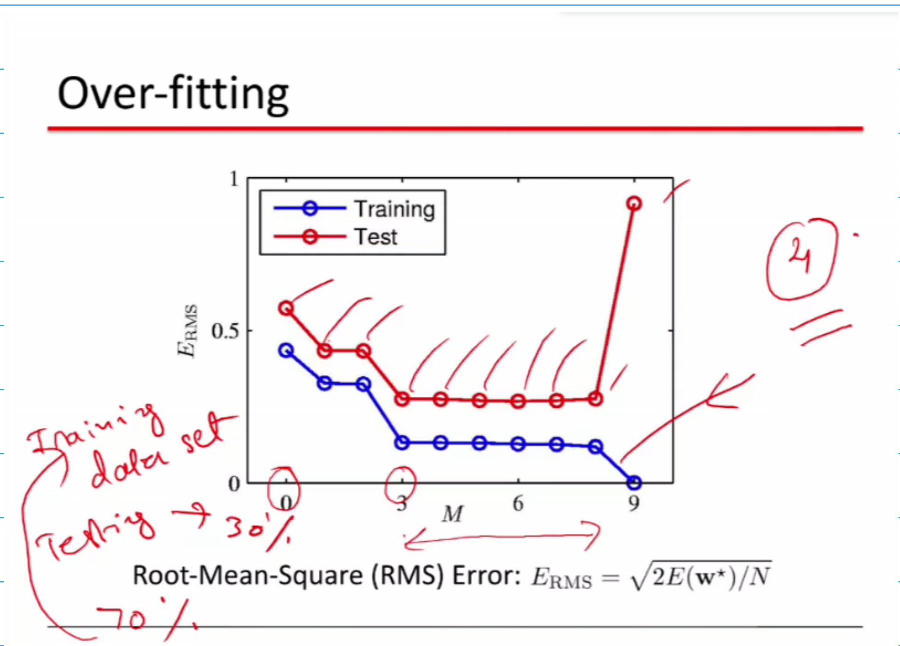
So model won't overfit

$$\left.\begin{array}{l} \min E(\omega) \\ \|\omega\| \le \epsilon \end{array}\right\}$$

~~wen~~ when you curtail the freedom of regression this way it's called Lasso regression //

$$\left.\begin{array}{l} \min E(\omega) \\ \omega^2 \le E \end{array}\right\}$$

ridge regression //



it's mostly because of larger coefficients

larger degree could give larger coefficients

	$M=0$	$M=1$	$M=3$	$M=9$
$w_0^\star$	0.19	0.82	0.31	0.35
$w_1^\star$		-1.27	7.99	232.37
$w_2^\star$			-25.43	-5321.83
$w_3^\star$			17.37	48568.31
$w_4^\star$				-231639.30
$w_5^\star$				640042.26
$w_6^\star$				-1061800.52
$w_7^\star$				1042400.18
$w_8^\star$				-557682.99
$w_9^\star$				125201.43

min $E(w)$ subject to $w_1^2 + w_2^2 + \cdots w_9^2 \le \eta$

This is a constrained optimization problem

To make it unconstrained

$$\min \; E(w) + \lambda(w_1^2 + w_2^2 + \cdots + w_9^2)$$

$\lambda \ge 0$ [maybe lagrange multipliers]

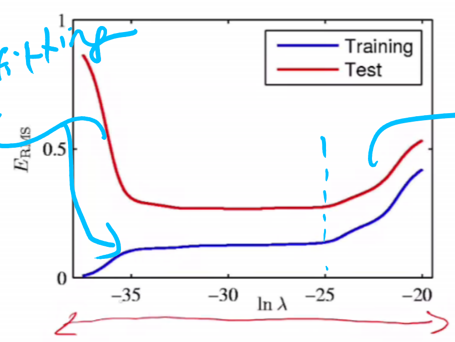
λ as large value will keep weights low but we need to minimize $E(w)$ as well!

λ too small $\Rightarrow$ Overfitting happens!!

λ is balancing factor between minimizing growth of w and $E(w)$.

Regularization: E_{RMS} vs. $\ln \lambda$

overfitting



Lesser importance to minimising $E(w)$

More imp to min $\sum w^2$

How can we incorporate Bayesian