WEEK 5.2

REGULARIZATION

→ is a general approach to add a 'complexity parameter' to a learning algo. in order to penalize complexity in models (meaning that we prefer learning a simple model than a complex one)

→ To be able to use this approach, the model parameters have to be continuous.
(so regression is fine, but DTs won't work due to discontinuous param.)

→ Looking into polynomial regression, if we penalize polynomials that have large coefficients, we'll get less wiggly solns.

↳ So we change the error measure tht. we use

$$\tilde{E}(w) = |y - \Phi w|^2 + \lambda |w|^2$$

adds a multiple λ of the modulus of the model parameters squared

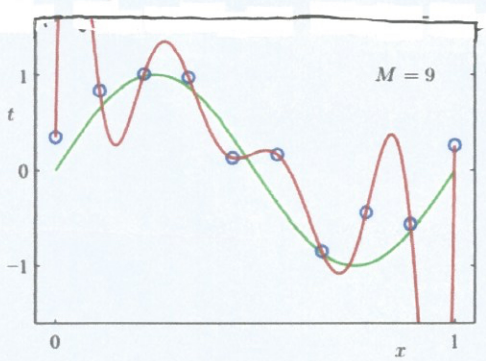
The soln will now become:

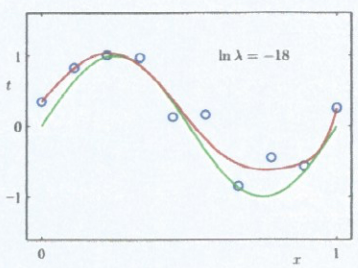
$$\hat{w} = (\Phi^T \Phi + \lambda l)^{-1} \Phi^T y$$

→ Pseudo-inverse soln, similar to linear regression

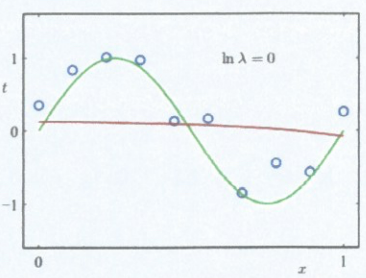
RIDGE REGRESSION

↳ Hence rather than using a discrete control parameter (like degree of polynomial), we can use a continuous parameter λ

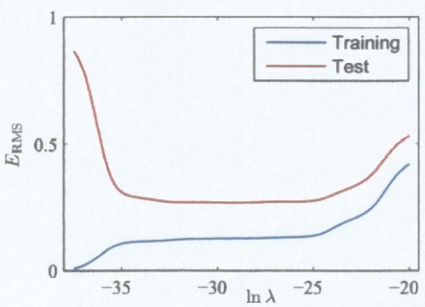




→ λ can be used to avoid overfitting



→ But can cause underfitting too if ln λ gets higher



	Standard Linear regression	Ridge regression
1) Task	Regression	Regression
2) Model structure	Linear regression model	Linear regression model
3) Score fn	Squared error (likelihood)	squared error w/ quadratic regularization
4) Optimization/ Search method	Calculus (analytical soln)	Calculus

↳ Notice how you can train the same model struct. w/ diff. score fns

HOW DO WE SET λ ?

→ It won't work if we set it based on the training set.

↳ WHY? Look at the graph on prev. pg. We would set $\ln\lambda$ to zero but that means we would do badly on the testing set.

→ If we want to optimise the setting of a complexity param., i.e. λ in ridge regression, we could use a validation set & cross-validation.

↓

1) Split labelled data to a training, validation and test set.
↳ i.e. 60% training, 20% validation, 20% test
2) Training set — use for training
3) Validation set — use to tune the 'control param.' according to performance on validation set → Best validation error
4) Testing set — to check how the final model performs

→ To set a continuous control param. like λ, we can pick a grid of values to search.

In practice, values are selected geometrically. → rather than linearly

e.g. Try $\lambda \in \{0.01, 0.1, 0.5, 1.0, 5.0, 10.0\}$