III. Polynomial Regression

$$y_i = -0.5 + 2x_i + 2x_i^2 + \epsilon_i$$

Now we try $y = w_0 + w_1 x + w_2 x^2$ and minimize the square loss

$$L(w_0, w_1, w_2) = \sum_{i=1}^{n} \left(y_i - w_0 - w_1 x_i - w_2 x_i^2 \right)^2$$

$$L(\underset{\sim}{w}) = \| \underset{\sim}{y} - \underset{\approx}{X} \underset{\sim}{w} \|^2$$

where $\underset{\sim}{w} = (w_0, w_1, w_2)^T \quad (3 \times 1)$

$$\underset{\approx}{X} = \begin{bmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ \vdots & \vdots & \vdots \\ 1 & x_n & x_n^2 \end{bmatrix} \quad (n \times 3)$$

$\underline{y} = (y_1 \dots y_n)^T \quad (n \times 3)$

another linear system

$$\underline{\underline{X}}^T \underline{\underline{X}} \, \underline{w} = \underline{\underline{X}}^T \underline{y}$$

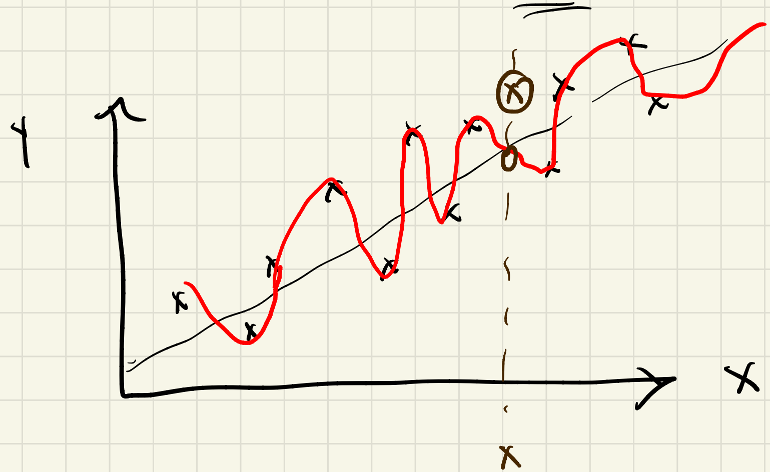
solve for $\underline{w}$

linear in $\underset{\sim}{w}$

Example 2 : what if we used a higher degree polynomial?

$$\underline{\underline{X}} = \begin{bmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^p \\ 1 & x_2 & x_2^2 & \cdots & x_2^p \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_n & x_n^2 & \cdots & x_n^p \end{bmatrix} \qquad (n \times p)$$

$$\underline{w} = (w_0, w_1, \ldots, w_p)^T \qquad ((p+1) \times 1)^T$$



IV The Generalized Linear Model

$$y(\underset{\sim}{x};\underset{\sim}{w}) = \sum_{j=1}^{m} w_j \, \phi_j(\underset{\sim}{x}) = \underset{\sim}{w}^T \underset{\sim}{\phi}(\underset{\sim}{x})$$

weights $\underline{w}^T = (w_0, w_1, \dots w_m)$

$\underset{\sim}{\phi} = (\phi_1, \dots, \phi_m)^T$ arbitrary basis functions

model is linear in $\underset{\sim}{w}$, but the basis functions $\phi(\underset{\sim}{x})$ can be non-linear

eg) our polynomial model: $\phi_1(x) = 1 \quad \phi_2(x) = x \quad \phi_3(x) = x^2$

eg) multivariate linear regression $\underset{\sim}{x} = (x_1 \dots x_d)$

$$y = \underline{w_0} + w_1 x_1 + \dots w_d x_d$$

is also a generalized linear model where
$\phi_1(x) = 1$
$\phi_2(x) = x_1$
$\phi_3(x) = x_2$

eg) multi-dimensional polynomials

fixed parameters

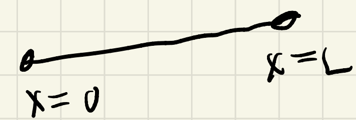
$$\phi_j(\tilde{x}) = \sum_{\alpha} \beta_\alpha \tilde{x}^{\alpha}$$

eg) radial basis functions

$$\phi_j(\tilde{x}) = \exp\left\{ - \frac{\| \tilde{x} - \tilde{x}_j \|^2}{2\ell^2} \right\}$$

eg) fourier series

$$\phi_{2j}(x) = \cos\left(\frac{2\pi j}{L} x \right)$$



$$\phi_{2j+1}(x) = \sin\left(\frac{2\pi j}{L} x \right)$$

How to fit the generalized model using least squares:

quadratic loss function

$$\mathcal{L}(\underset{\sim}{w}) = \sum_{i=1}^{N} \left[y(\underset{\sim}{x}_i; w) - y_i \right]^2$$

$$= \| \underline{\underline{\Phi}}\, \underline{w} - \underline{y} \|^2$$

$$= (\underline{\underline{\Phi}}\, \underline{w} - \underline{y})^T (\underline{\underline{\Phi}}\, \underline{w} - \underline{y})$$

$\underline{\underline{\Phi}}$ = design matrix

$\underline{\underline{\Phi}} \in \mathbb{R}^{n \times m}$

n = # observations

m = # basis functions

$\Phi_{ij} = \phi_j(\underset{\sim}{x}_i)$

$\underline{w}^T = (w_1, \ldots, w_m)$

To minimize the loss:

① Take $\nabla_{\underline{w}} \mathcal{L}(\underline{w})$, set $= 0$, solve for $\underline{w}$

solution: $(\underline{\underline{\Phi}}^T \underline{\underline{\Phi}})\,\underline{w} = (\underline{\underline{\Phi}}^T)\,\underline{y}$ as before

solve w/ numpy.linalg.lstsq

provide $\underline{\underline{\Phi}}$, $\underline{y}$ and obtain $\underline{w}$

IV. Measures of Predictive Accuracy

(1) Training and test datasets

- not a good idea to test how good your model is using the training dataset
- need to test model on unseen data - a test data set
- often, take your data and split (eg 70% train, 30% test)
- can plot predictions vs. observations

(2) Mean-squared error - scalar measure of goodness of fit

$$MSE = \frac{1}{n_{test}} \sum_{i=1}^{n_{test}} \left[y_{test,i} - w^T \phi(x_{t,i}) \right]^2$$

hard to understand the absolute meaning of MSE, so we can instead use relative MSE (RMSE)

$$\text{RMSE} = \frac{\text{MSE of your model}}{\text{MSE of simplest possible model}}$$

simplest: $y_{simplest} = \hat{u} = \frac{1}{n} \sum_{i=1}^{n} y_i$

if RMSE < 1 $\rightarrow$ you are doing better than simplest possible model

RMSE > 1 $\rightarrow$ the simplest model outperforms yours

(3) Coefficient of Determination R^2

$R^2 = 1 - \text{RMSE}$: tells us what % of the variance of the test data is explained by your model