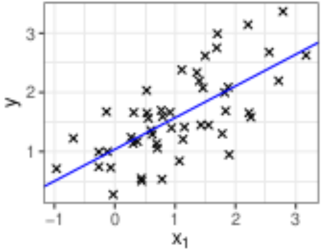
LINEAR REGRESSION

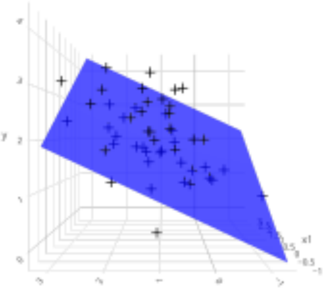
- Idea: predict $y \in \mathbb{R}$ as **linear** combination of features[1]:

$$\hat{y} = f(\mathbf{x}) = \boldsymbol{\theta}^\top \mathbf{x} = \theta_0 + \theta_1 x_1 + \dots + \theta_p x_p$$

$\rightsquigarrow$ find loss-optimal params to describe relation $y|\mathbf{x}$

- Hypothesis space: $\mathcal{H} = \{f(\mathbf{x}) = \boldsymbol{\theta}^\top \mathbf{x} \mid \boldsymbol{\theta} \in \mathbb{R}^{p+1}\}$

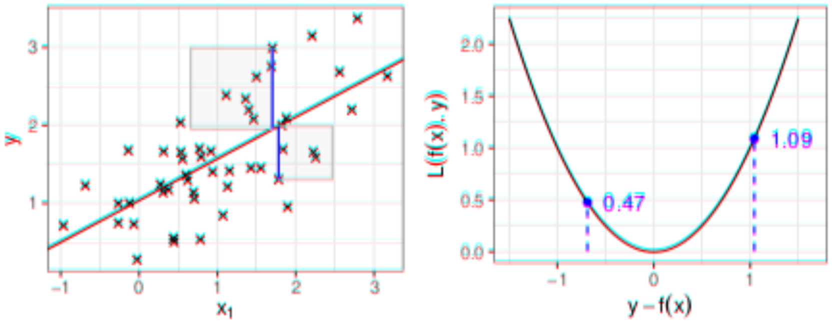




[1] Actually, special case of linear model, which is linear combo of *basis functions* of features $\rightsquigarrow$ Polynomial Regression Models

LOSS PLOTS

We will often visualize loss effects like this:



- Data as $y \sim x_1$
- Prediction hypersurface
  ⇝ here: line
- Residuals $r = y - f(\mathbf{x})$
  ⇝ squares to illustrate loss
- Loss as function of residuals
  ⇝ strength of penalty?
  ⇝ symmetric?
- Highlighted: loss for residuals shown on LHS

STATISTICAL PROPERTIES

- LM with $L2$ loss intimately related to classical stats LM
- Assumptions
  - $\mathbf{x}^{(i)}$ **iid** for $i \in \{1, \dots, n\}$
  - **Homoskedastic** (equivariant) **Gaussian** errors

$$\mathbf{y} = \mathbf{X}\boldsymbol{\theta} + \boldsymbol{\epsilon}, \ \boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \boldsymbol{I})$$

    ⇝ y conditionally independent & normal: $\mathbf{y} | \mathbf{X} \sim \mathcal{N}(\mathbf{X}\boldsymbol{\theta}, \sigma^2 \boldsymbol{I})$
  - Uncorrelated features

    ⇝ multicollinearity destabilizes effect estimation
- If assumptions hold: statistical **inference** applicable
  - Hypothesis tests on significance of effects, incl. p-values
  - Confidence & prediction intervals via student-t distribution
  - Goodness-of-fit measure $R^2 = 1 - \text{SSE} \ / \ \underbrace{\text{SST}}_{\sum_{i=1}^{n}(y^{(i)} - \bar{y})^2}$

    ⇝ SSE = part of data variance *not* explained by model