Open questions

- How do I quantify the measurement noise?

  ↳ by maximizing the likelihood

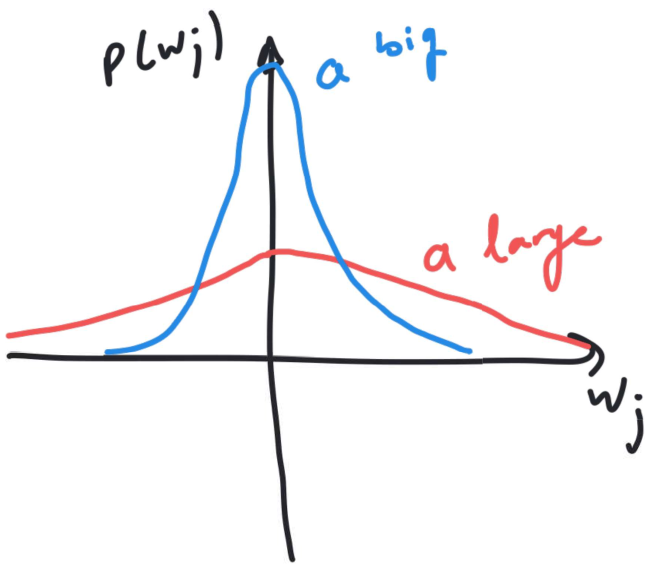
- How do we avoid overfitting?

  ↳ maximize the posterior but don't use uniform distribution for prior on weights

- How do I quantify epistemic uncertainty induced by limited data?
- How do I choose any remaining parameters?
- How do I choose which basis functions to keep?

PREDICTIVE
SCIENCE LABORATORY

Gaussian prior on weights



$$w_j \sim N(0, a^{-1})$$

we can pick this

precision $= \dfrac{1}{\text{variance}}$

controls spread of the prior

$$p(\underline{w}) = \prod_{j=1}^{m} p(w_j)$$

$$\propto \exp\left\{-\frac{a}{2}\sum_{j=1}^{m} w_j^2\right\}$$

Maximum a posteriori estimate

posterior $\propto$ likelihood $\times$ prior

$$p(\underline{w} \mid x_{1:N}, y_{1:N}, \sigma) \propto p(y_{1:N} \mid x_{1:N}, \sigma^2)\, p(\underline{w})$$

$$\max_{w} \log post = \log like + \log p(w)$$

$$= -\frac{1}{2\sigma^2}\sum_{i=1}^{N}\left(y_i - \underline{\varphi}^T(x_i)\underline{w}\right)^2 - \frac{\alpha}{2}\sum_{j=1}^{m} w_j^2 + const.$$

change sign

$$\min_{\underline{w}} \underbrace{\sum_{i=1}^{N}\left(y_i - \underline{\varphi}^T(x_i)\underline{w}\right)^2}_{SE} + \underbrace{\frac{\alpha}{2}\sum_{j=1}^{m} w_j^2}_{\text{Regularization term}}$$

square error

regularization term (pulls weights closer to 0)

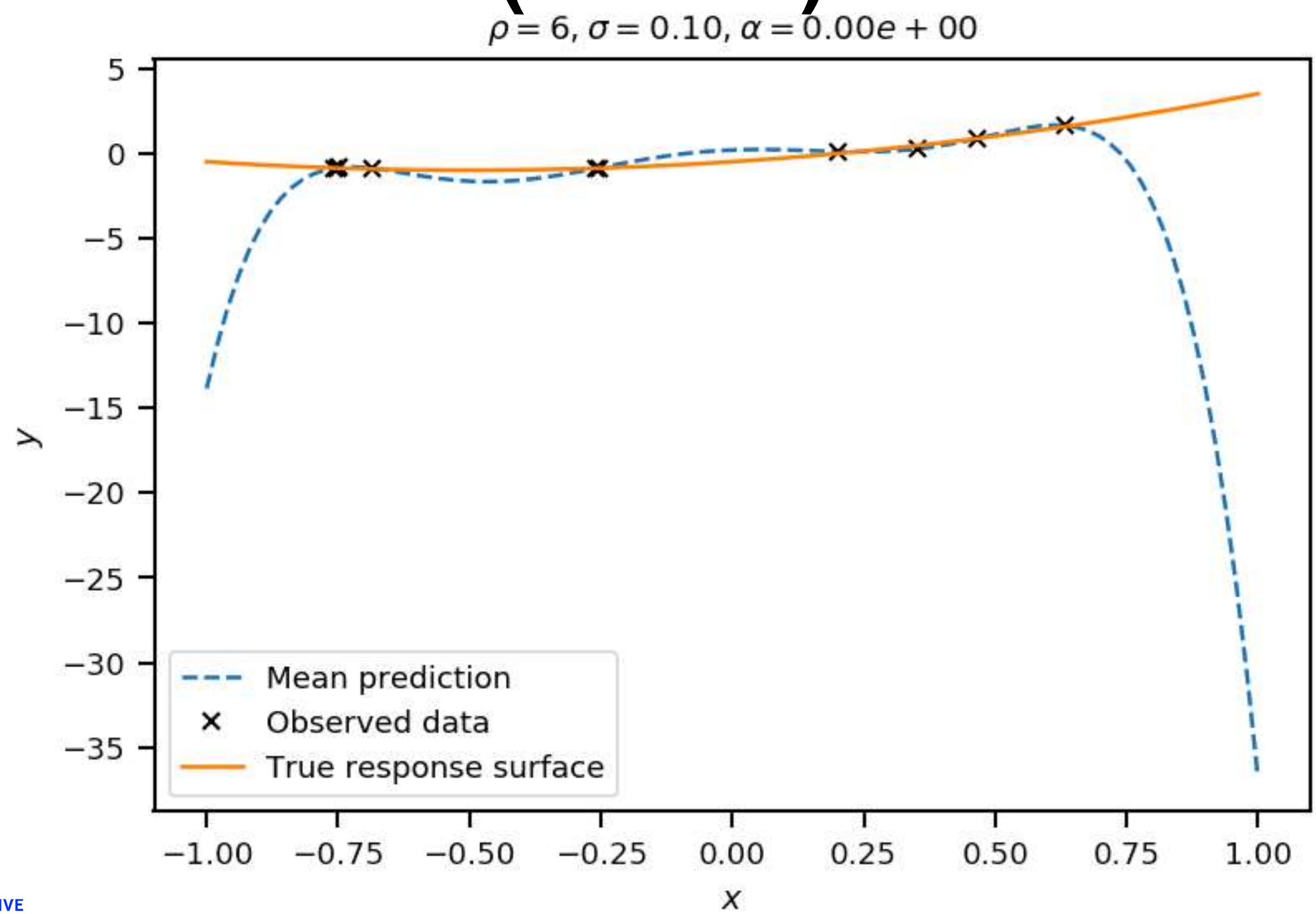
helps to avoid overfitting

PREDICTIVE SCIENCE LABORATORY

Example: Degree 6 polynomial

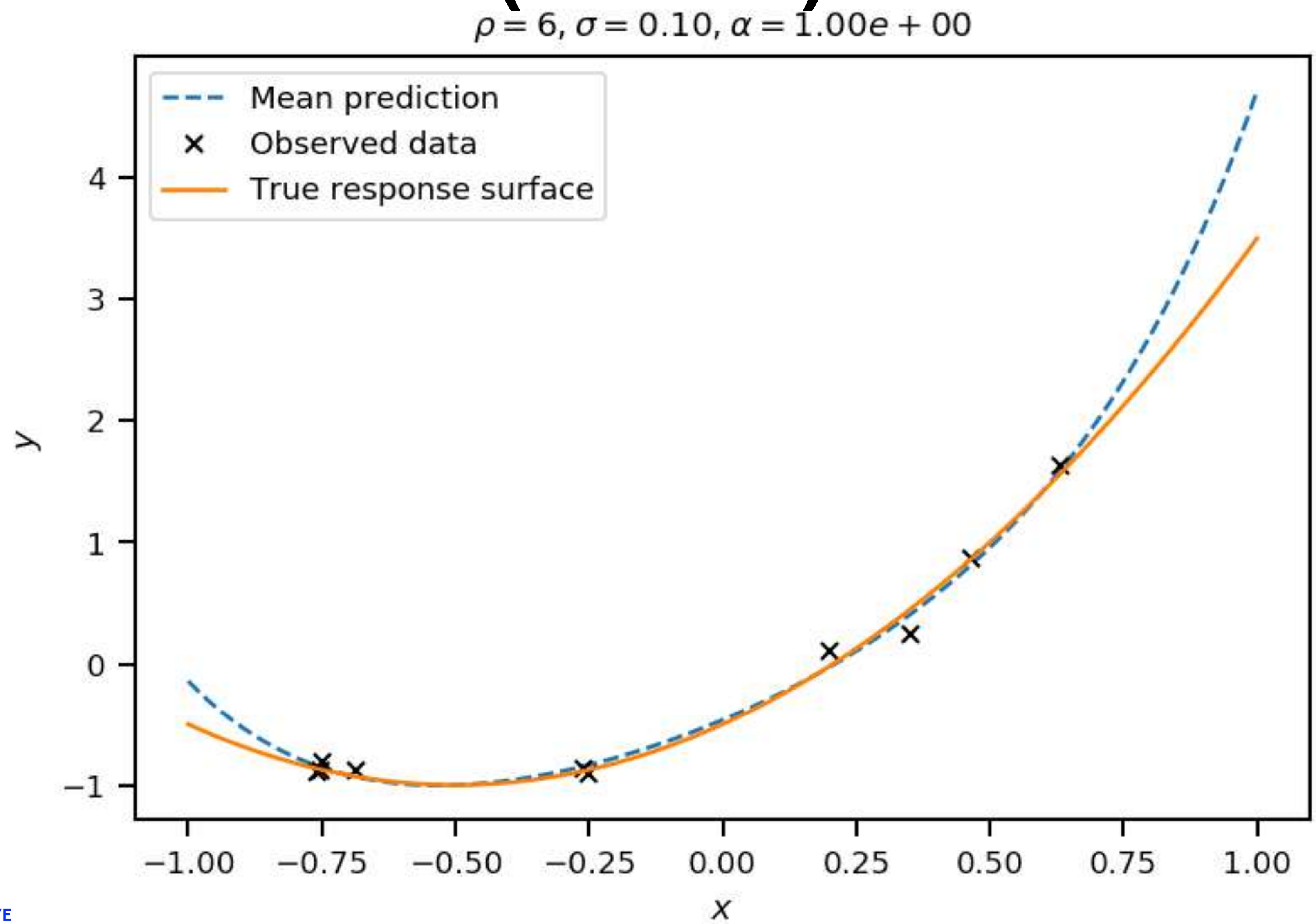
regularizer

$$(\alpha = 0)$$



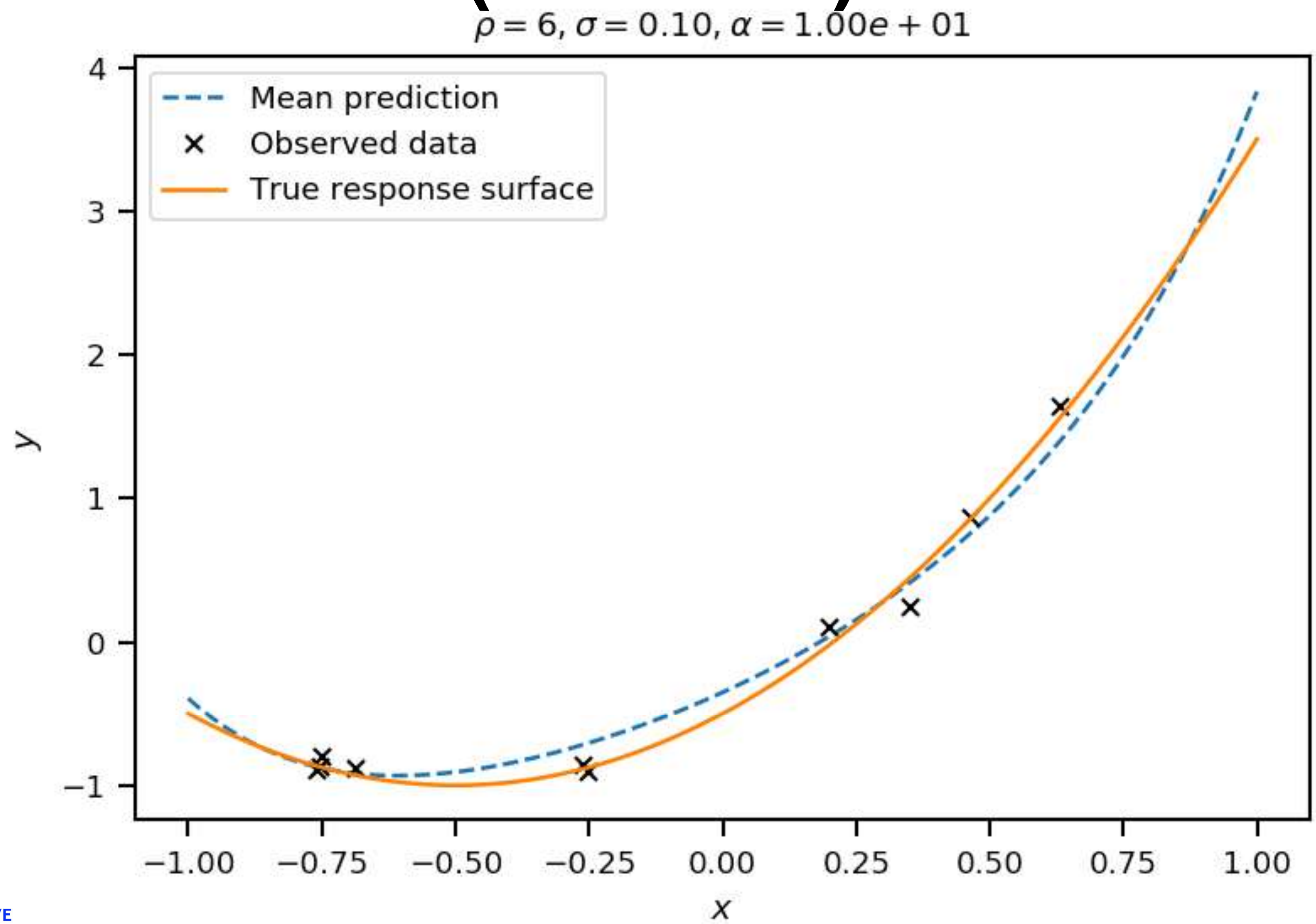
PREDICTIVE
SCIENCE LABORATORY

Example: Degree 6 polynomial $(\alpha = 1)$



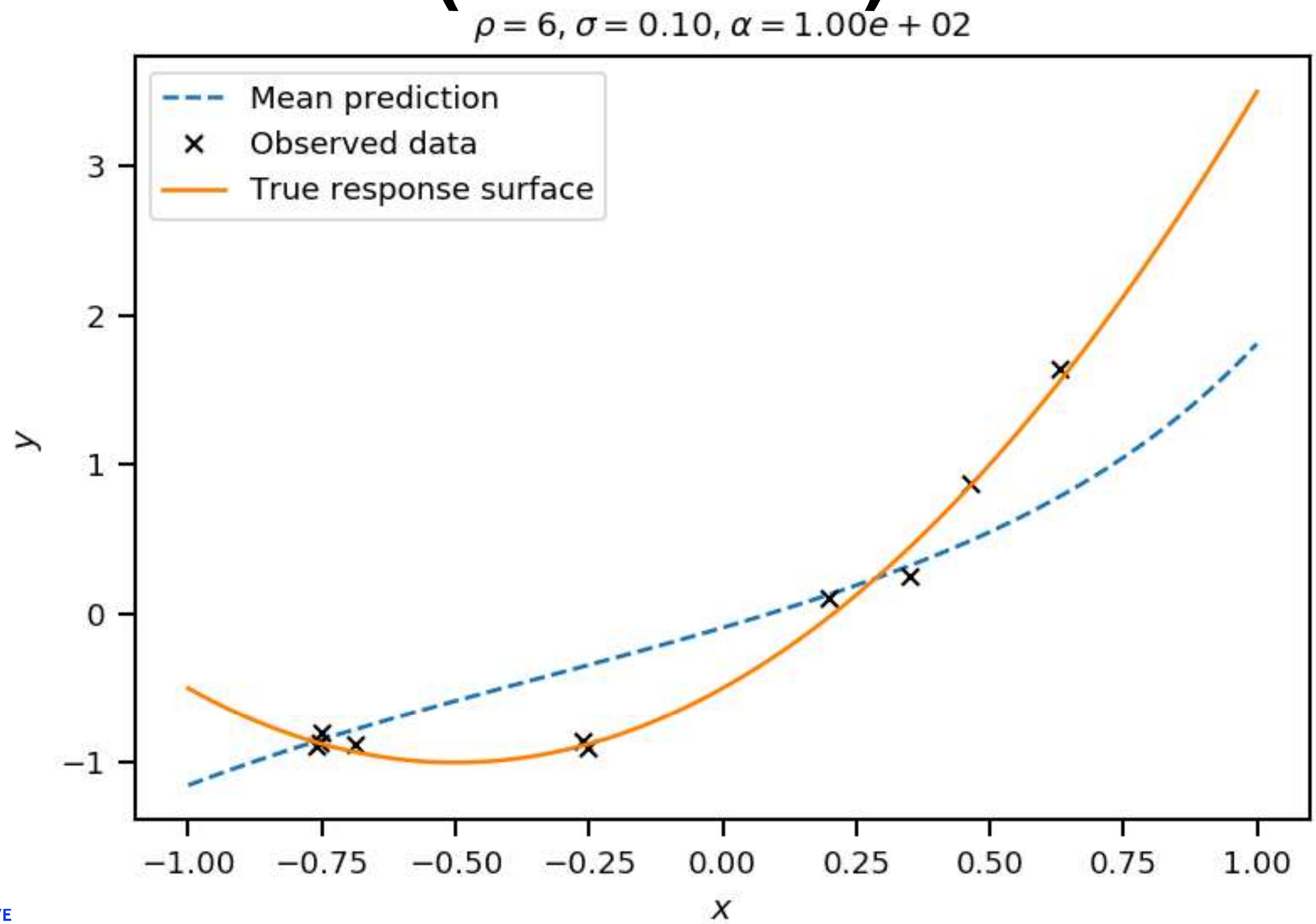
PREDICTIVE SCIENCE LABORATORY

Example: Degree 6 polynomial $(\alpha = 10)$



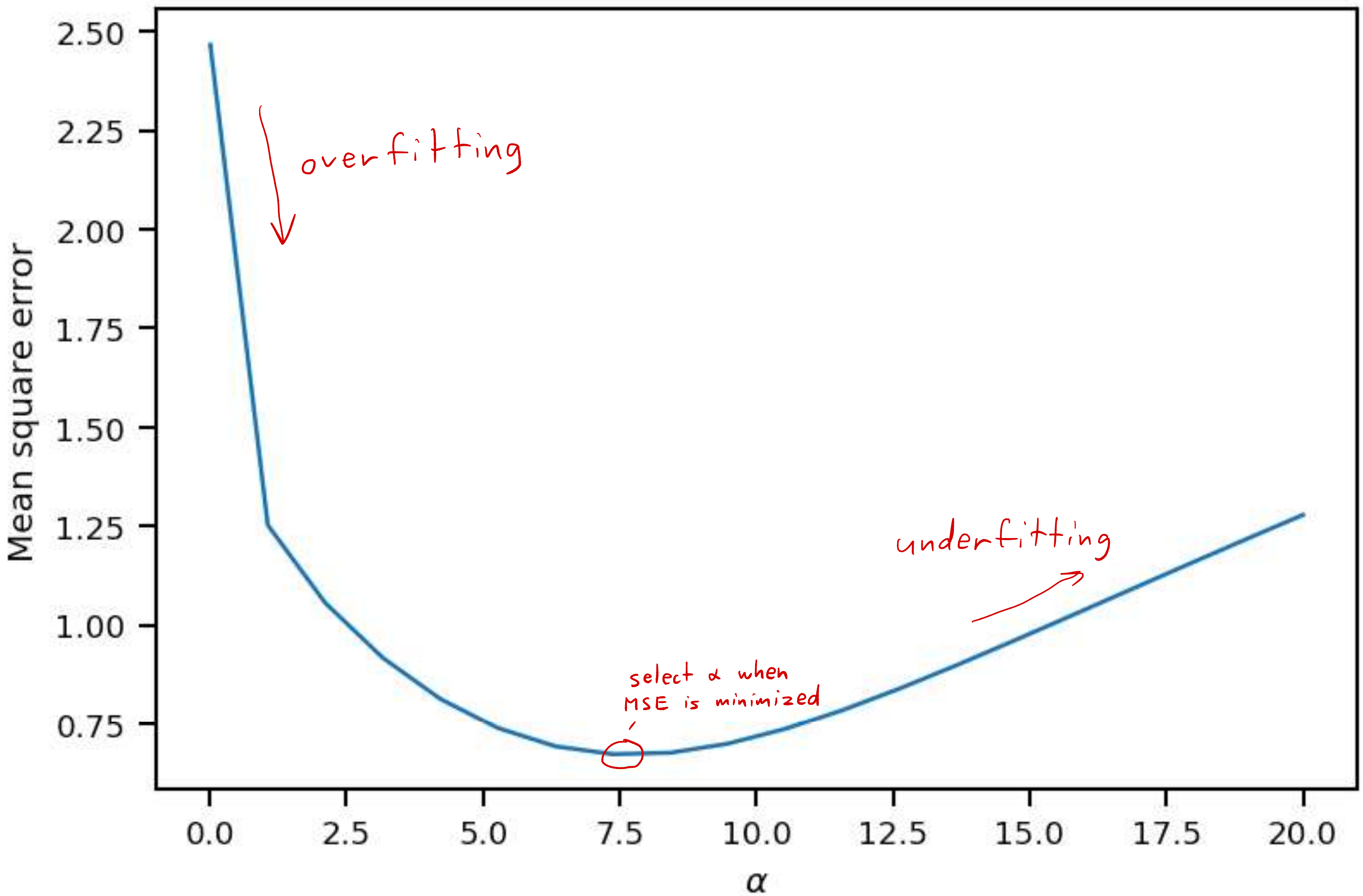
PREDICTIVE
SCIENCE LABORATORY

Example: Degree 6 polynomial $(\alpha = 100)$



PREDICTIVE
SCIENCE LABORATORY

Mean square error over a validation dataset as a function of α



PREDICTIVE
SCIENCE LABORATORY