Generalized Linear Regression and Bias-Variance Tradeoff

UNIVERSITY of WASHINGTON

Process

Collect a **data set**

Decide on a **model**

Find the function which fits the data best

Choose a loss function

Pick the function which minimizes loss on data

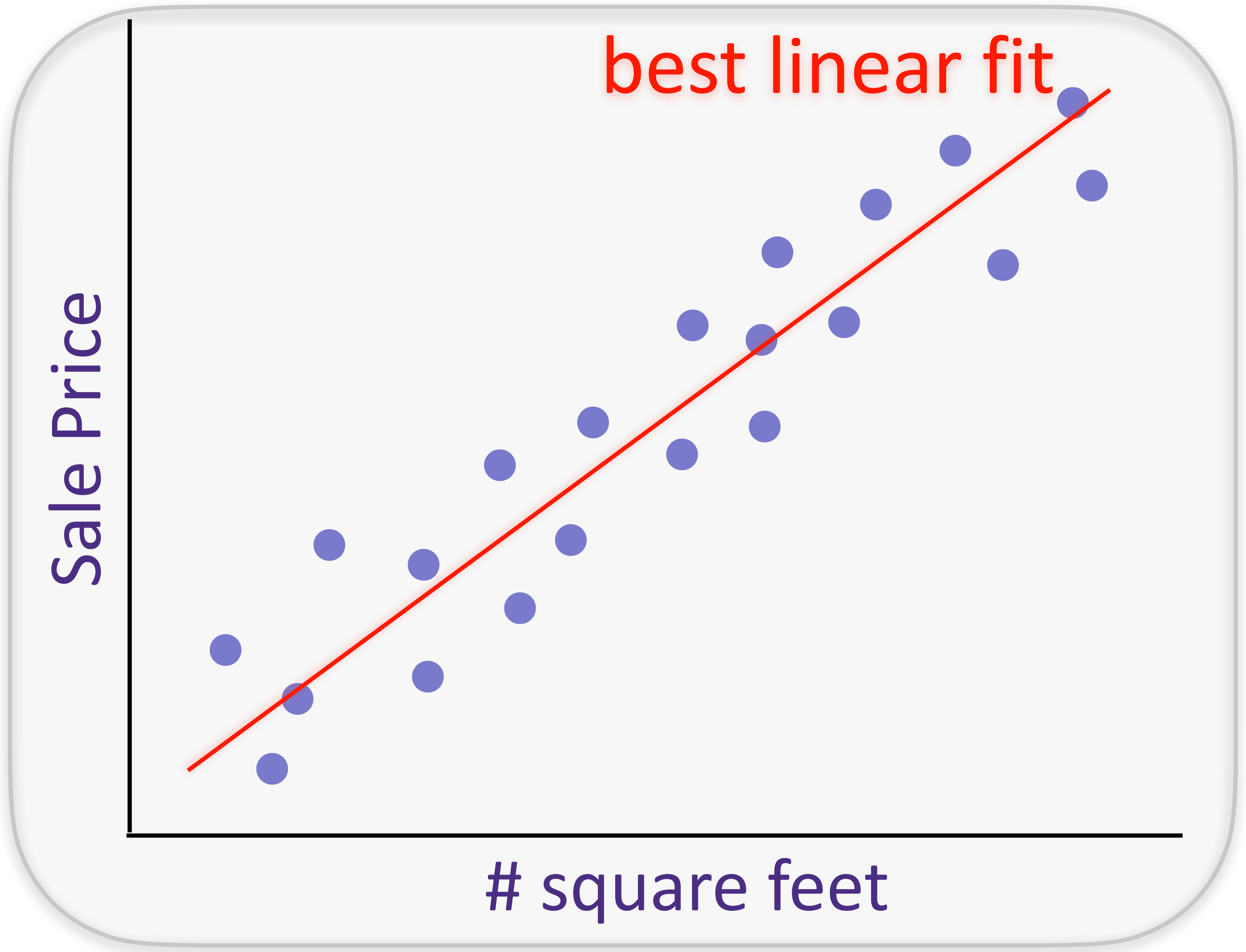
Use function to make prediction on new examples

The regression problem

Given past sales data on zillow.com, predict:

y = **House sale price** *from*

x = **{# sq. ft., zip code, date of sale, etc.}**



Training Data: $\{(x_i, y_i)\}_{i=1}^n$

$x_i \in \mathbb{R}^d$
$y_i \in \mathbb{R}$

Hypothesis:

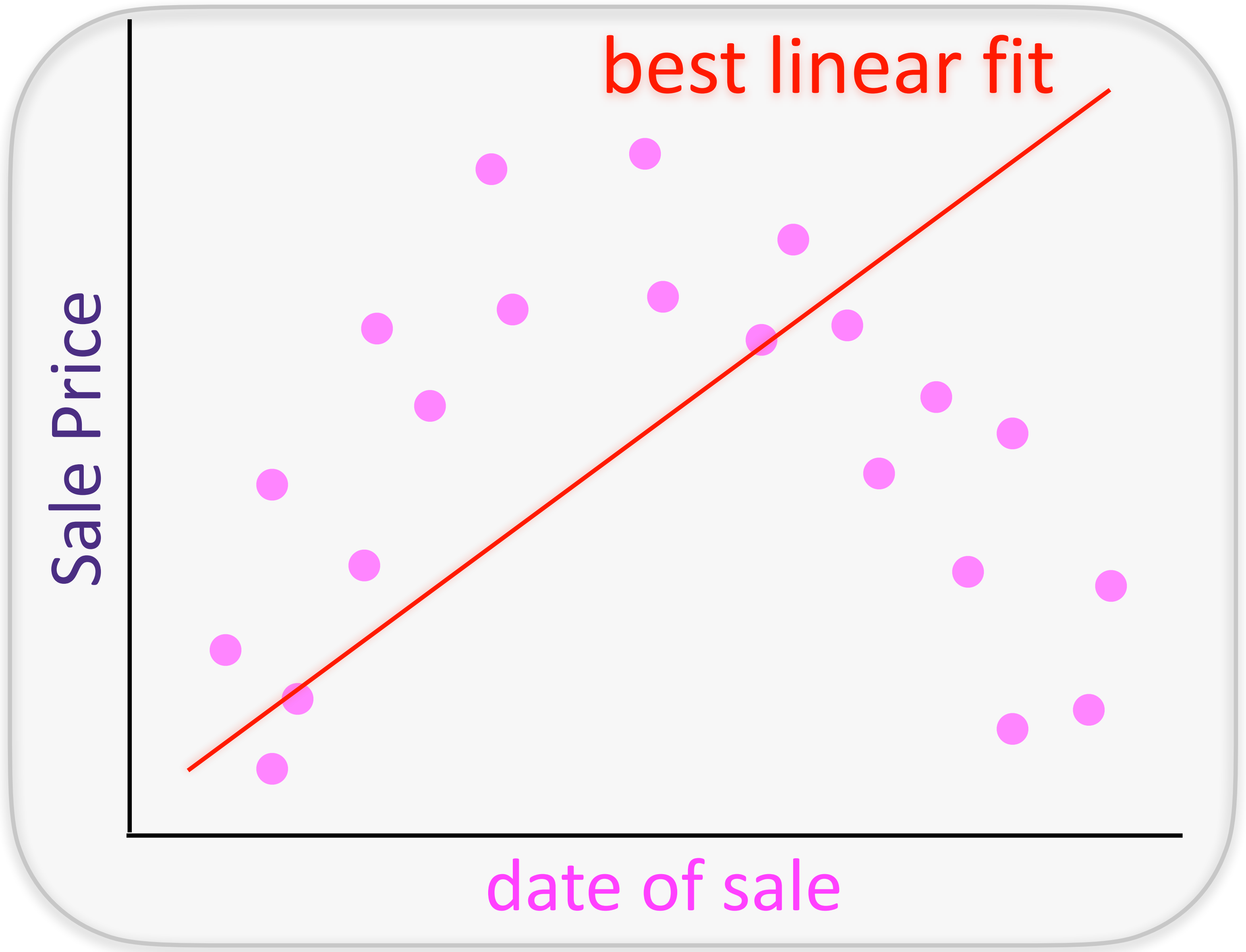
Loss:

The regression problem

Given past sales data on zillow.com, predict:

y = **House sale price** *from*

x = **{# sq. ft., zip code, date of sale, etc.}**

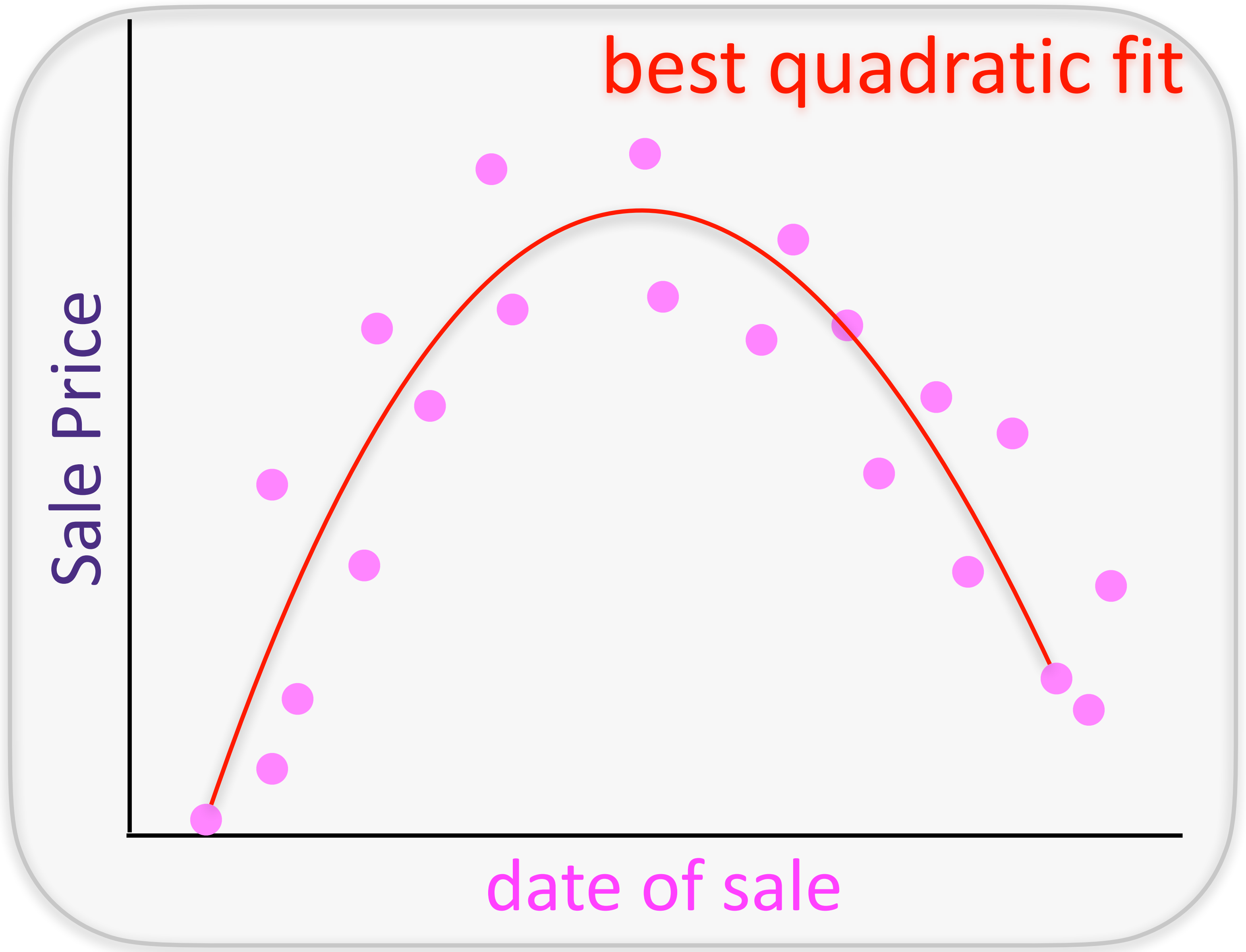


Quadratic Regression

Given past sales data on zillow.com, predict:

y = House sale price *from*

x = {# sq. ft., zip code, date of sale, etc.}



Training Data: $\{(x_i, y_i)\}_{i=1}^n$ $\quad x_i \in \mathbb{R}^d$, $y_i \in \mathbb{R}$

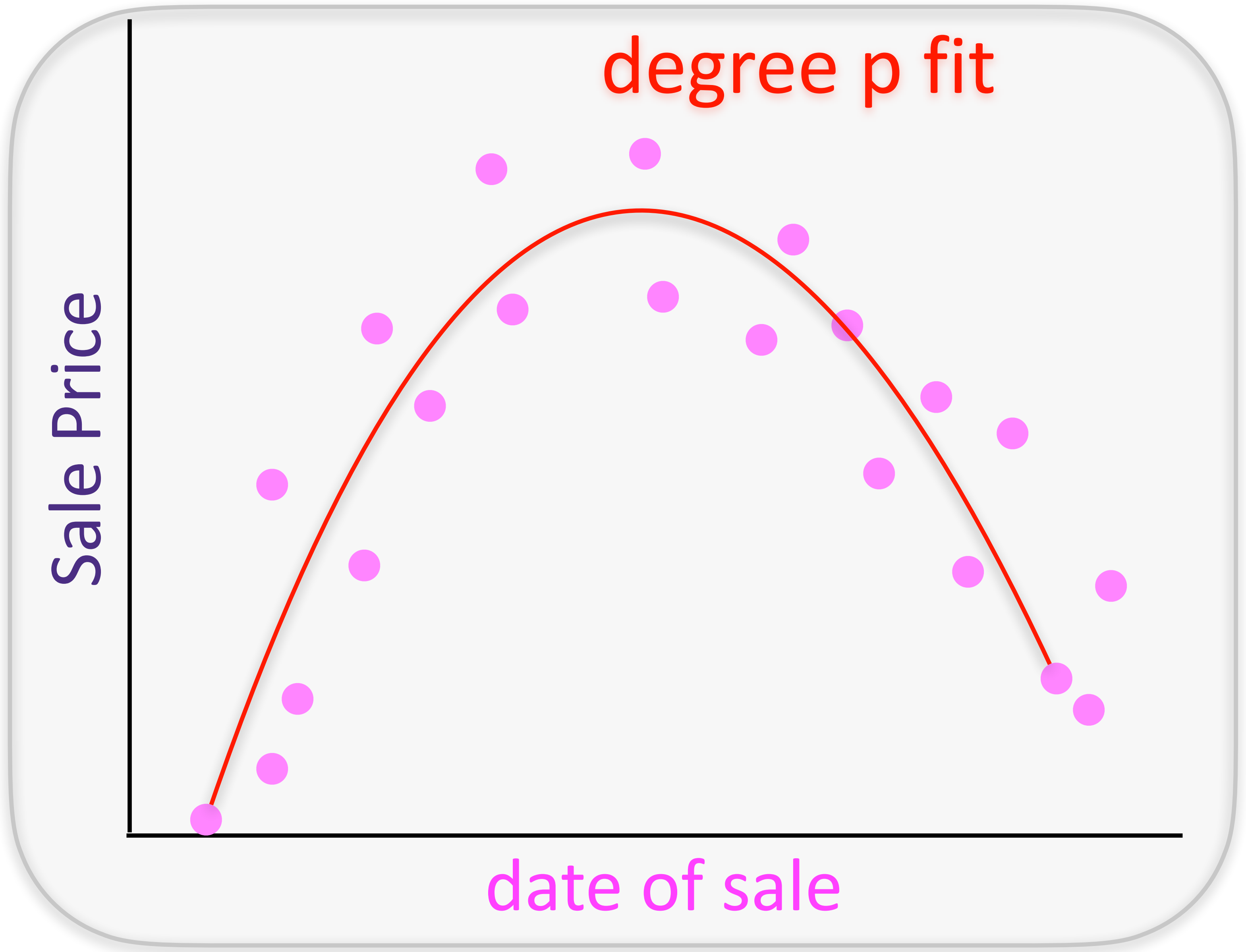
Hypothesis:

Polynomial regression

Given past sales data on zillow.com, predict:

y = **House sale price** *from*

x = **{# sq. ft., zip code, date of sale, etc.}**



Training Data: $\{(x_i, y_i)\}_{i=1}^n$ $\quad x_i \in \mathbb{R}^d$, $y_i \in \mathbb{R}$

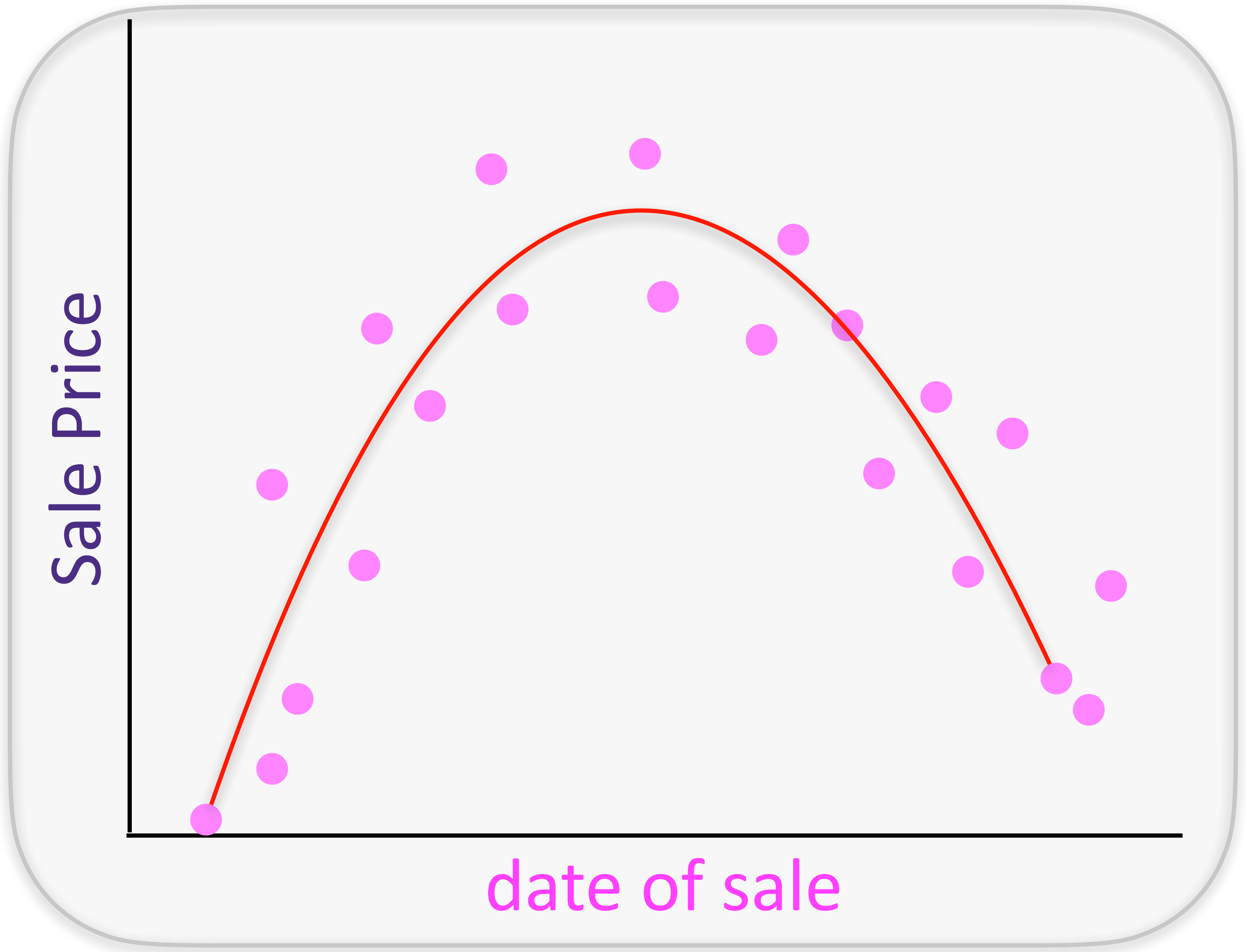
Hypothesis:

Generalized linear regression

Given past sales data on zillow.com, predict:

y = **House sale price** *from*

x = **{# sq. ft., zip code, date of sale, etc.}**



Training Data: $\{(x_i, y_i)\}_{i=1}^n$

$x_i \in \mathbb{R}^d$
$y_i \in \mathbb{R}$

Hypothesis:

Generalized Linear Regression

Training Data:

$$\{(x_i, y_i)\}_{i=1}^n$$

$x_i \in \mathbb{R}^d$
$y_i \in \mathbb{R}$

Transformed data:

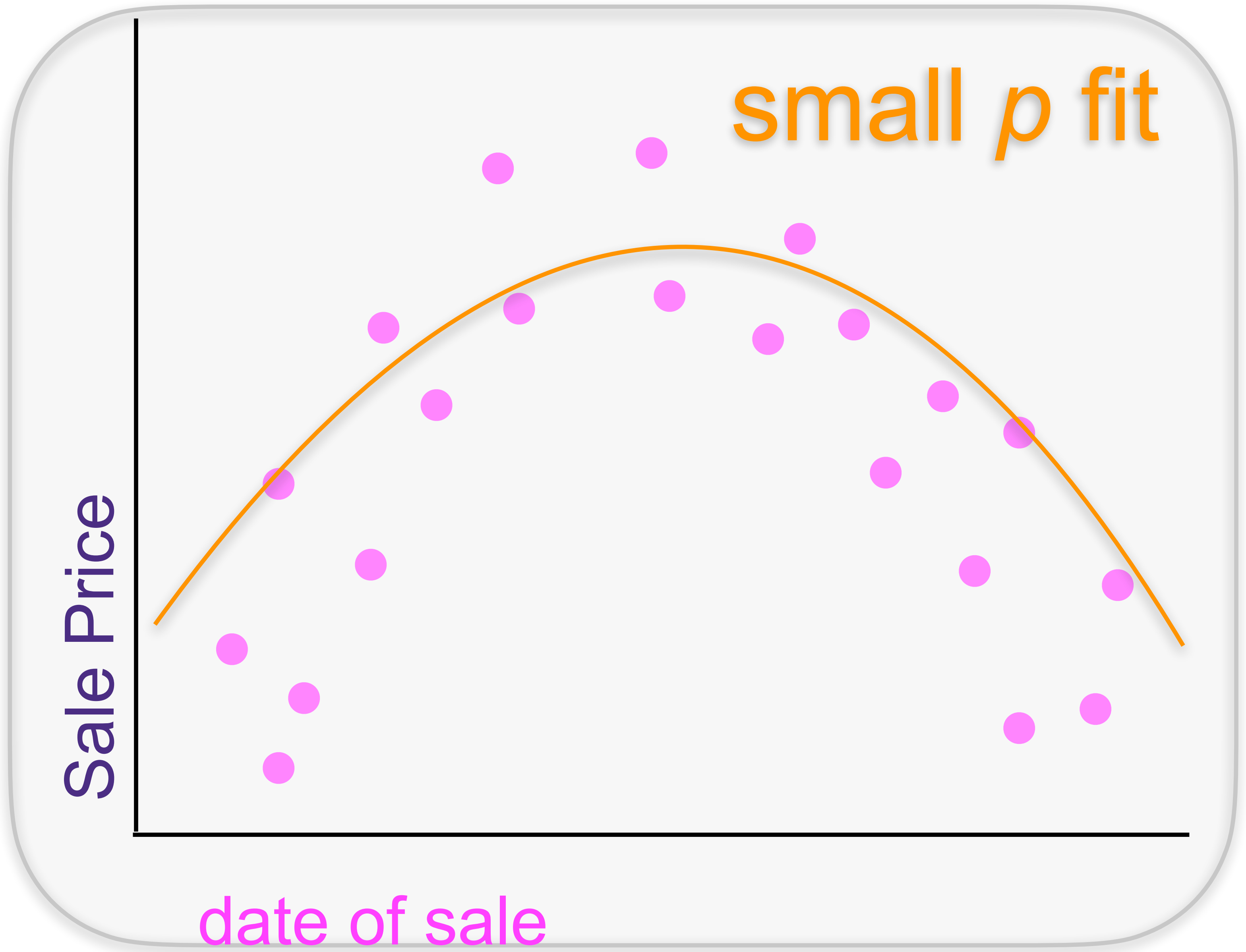
Hypothesis:

Loss:

The regression problem

Training Data:

$$\{(x_i, y_i)\}_{i=1}^n \qquad x_i \in \mathbb{R}^d \quad y_i \in \mathbb{R}$$



Transformed data:

$$h(x) = \begin{bmatrix} h_1(x) \\ h_2(x) \\ \vdots \\ h_p(x) \end{bmatrix}$$

Hypothesis: linear in h

$$y_i \approx h(x_i)^T w \quad w \in \mathbb{R}^p$$

Loss: least squares

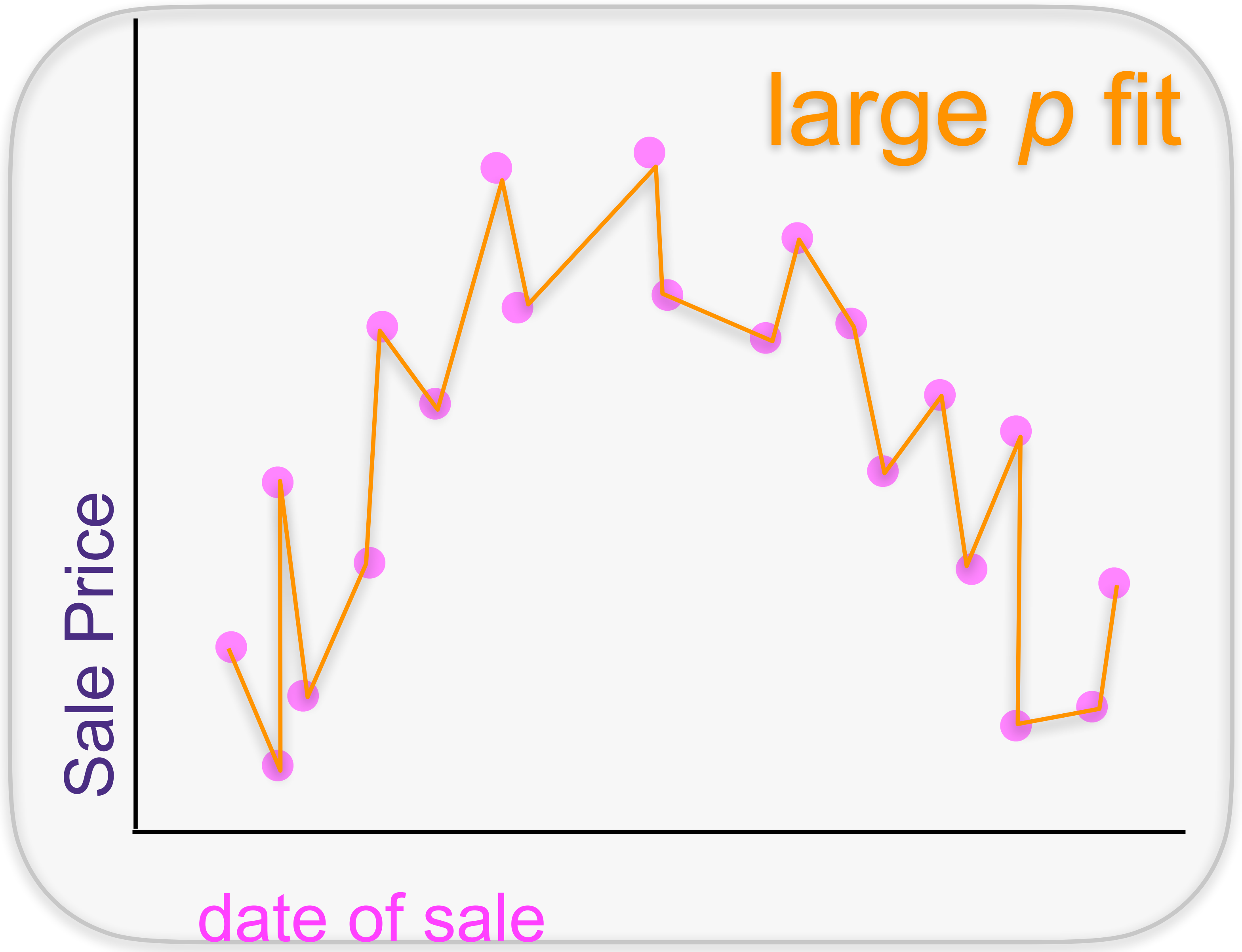
$$\min_w \sum_{i=1}^n \left(y_i - h(x_i)^T w\right)^2$$

The regression problem

Training Data:

$$\{(x_i, y_i)\}_{i=1}^{n}$$

$x_i \in \mathbb{R}^d$
$y_i \in \mathbb{R}$



Transformed data:

$$h(x) = \begin{bmatrix} h_1(x) \\ h_2(x) \\ \vdots \\ h_p(x) \end{bmatrix}$$

Hypothesis: linear in h

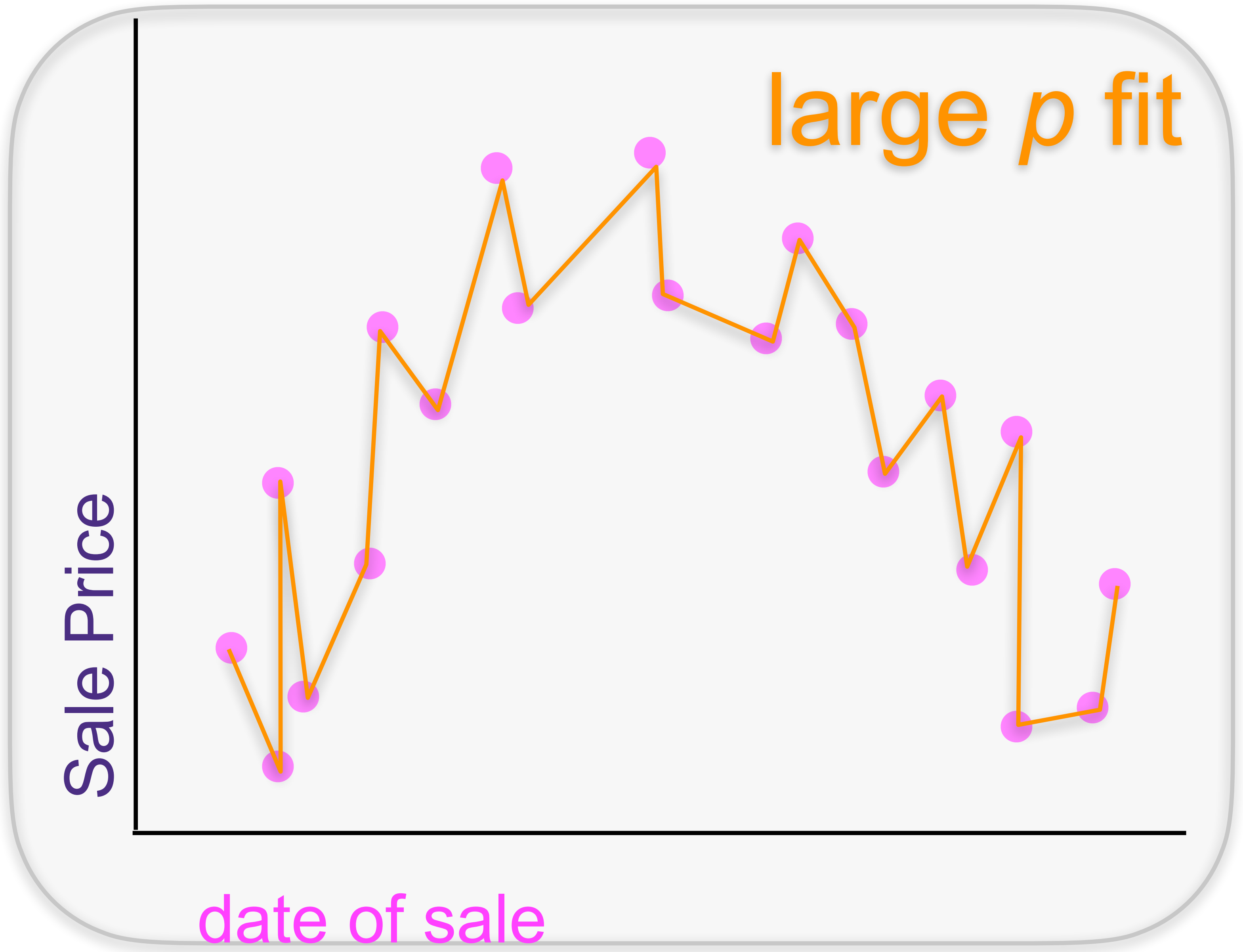
$$y_i \approx h(x_i)^T w \quad w \in \mathbb{R}^p$$

Loss: least squares

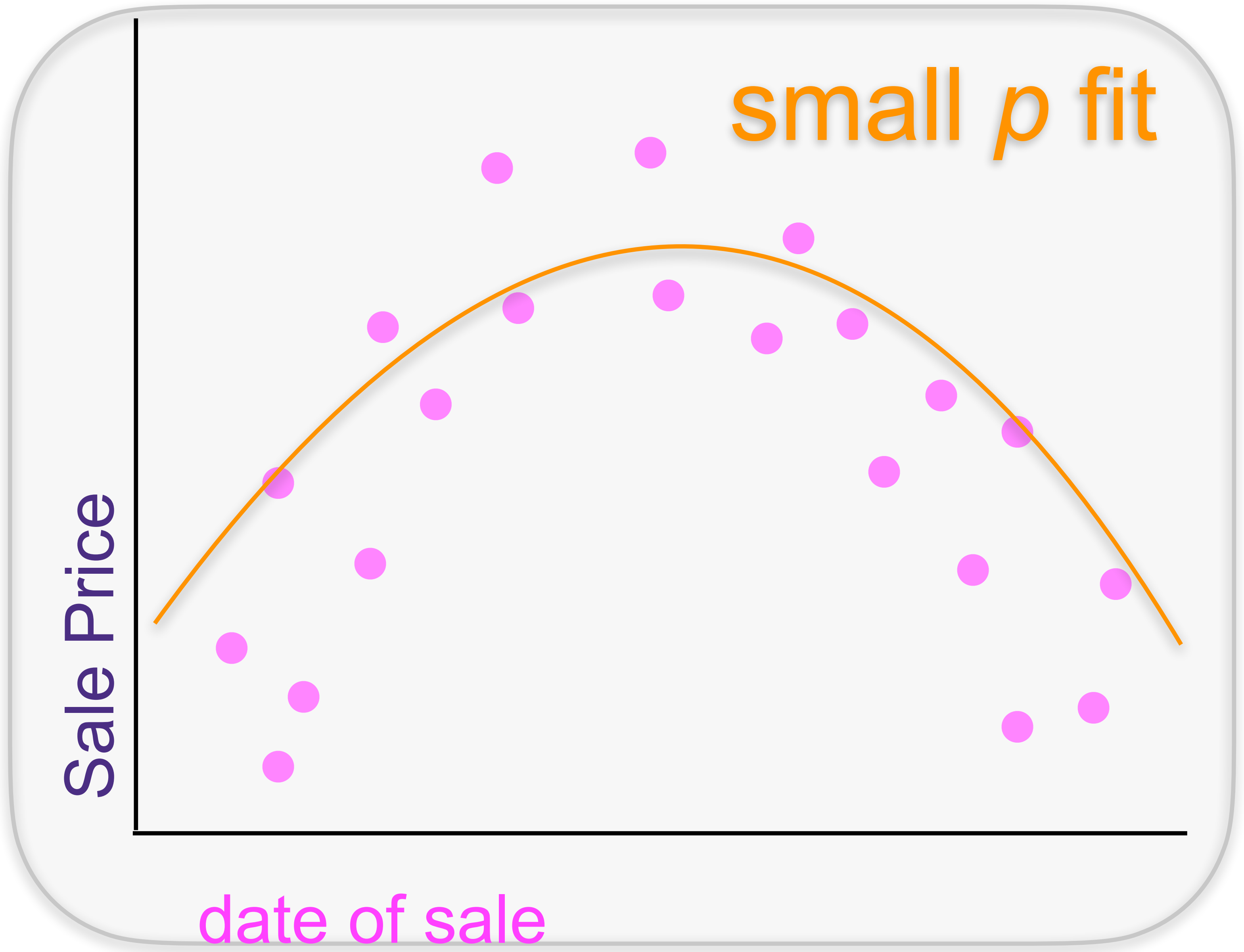
$$\min_w \sum_{i=1}^{n} \left(y_i - h(x_i)^T w\right)^2$$

Which is better?

A: large p



B: small p

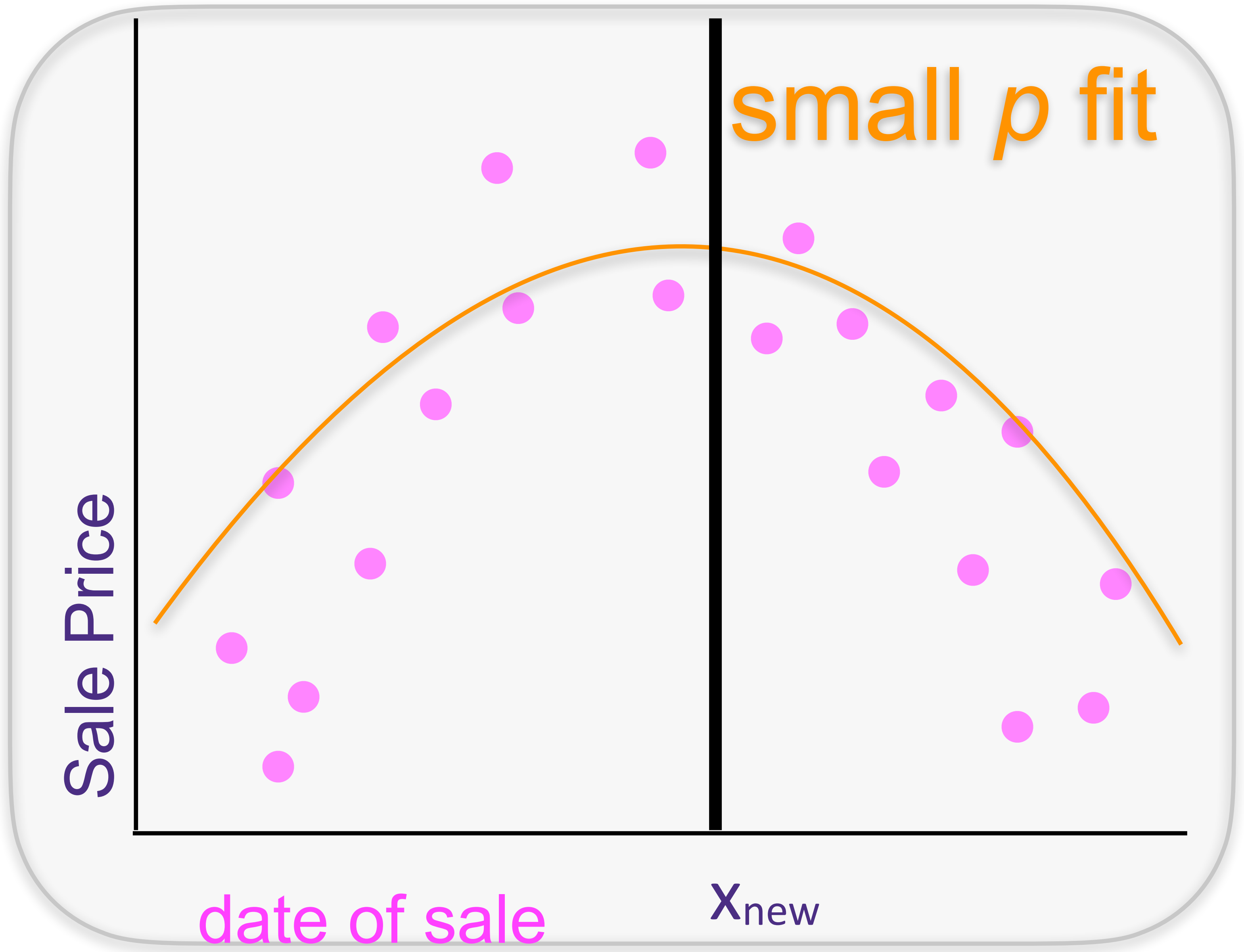


What is our goal, anyway?

Predicting sale price for a new house: A vs B

A: large p

B: small p

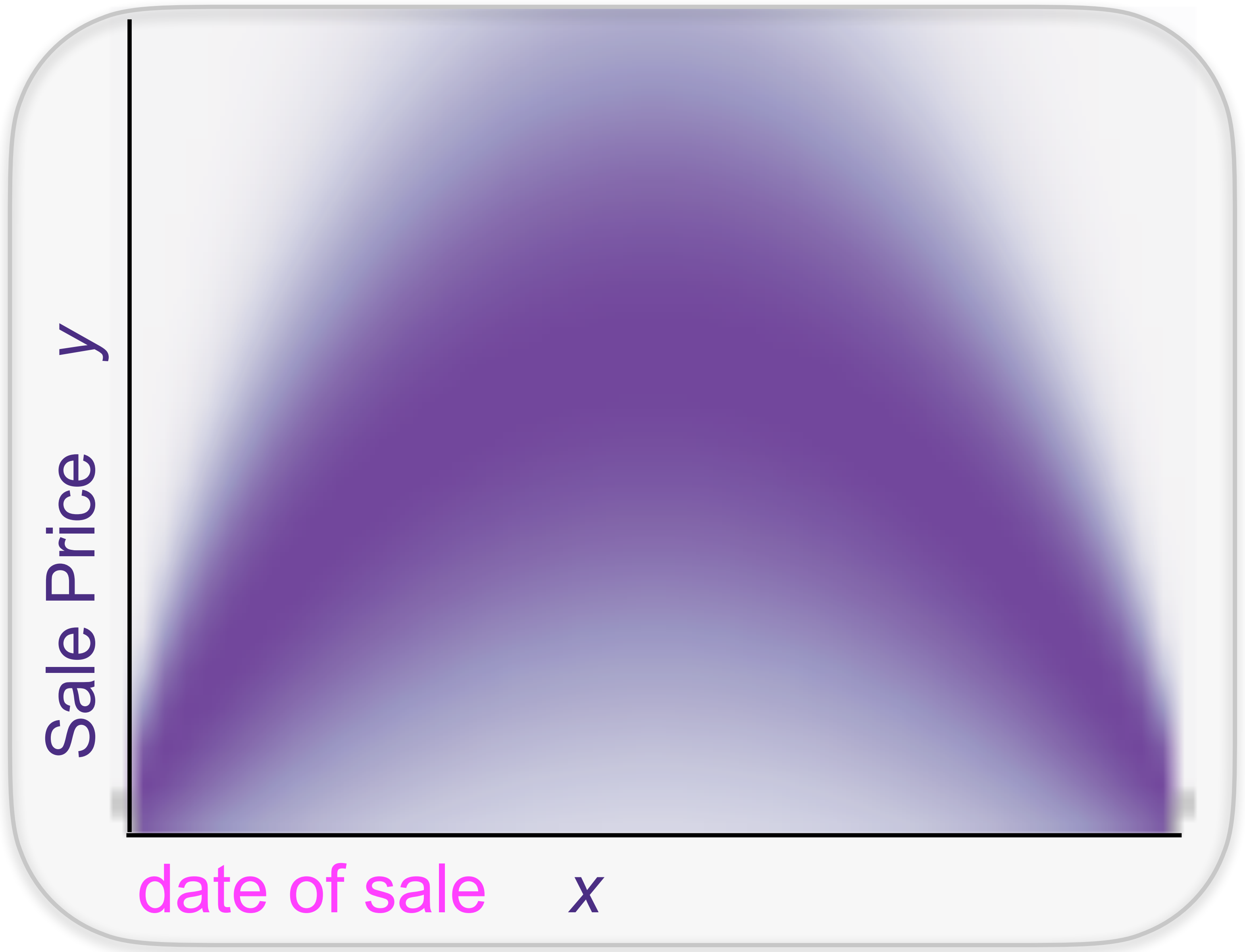


Our goal is to predict prices for new houses

that "look like" the houses in our training data

What do we mean by “look like”?

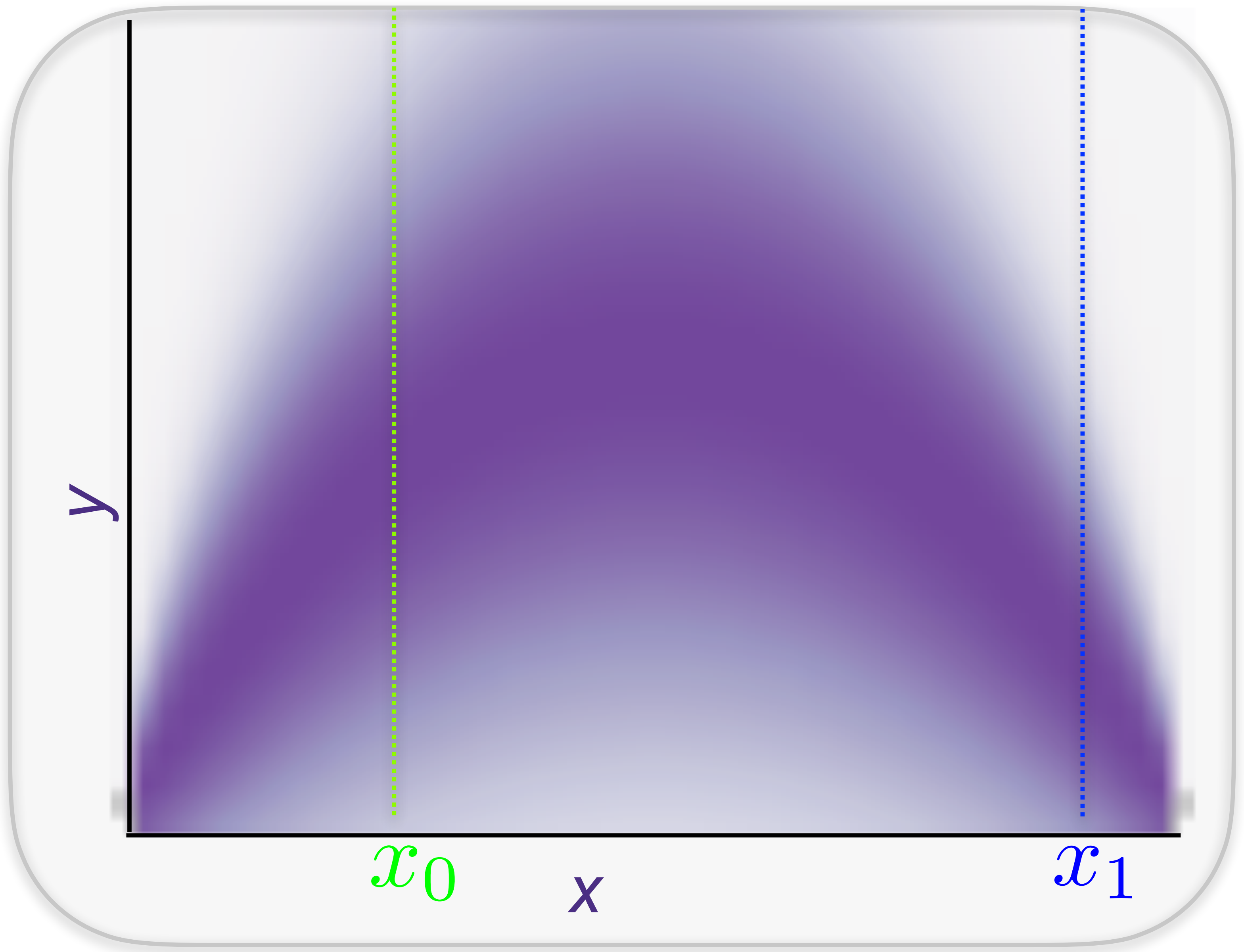
$$P_{XY}(X = x, Y = y)$$



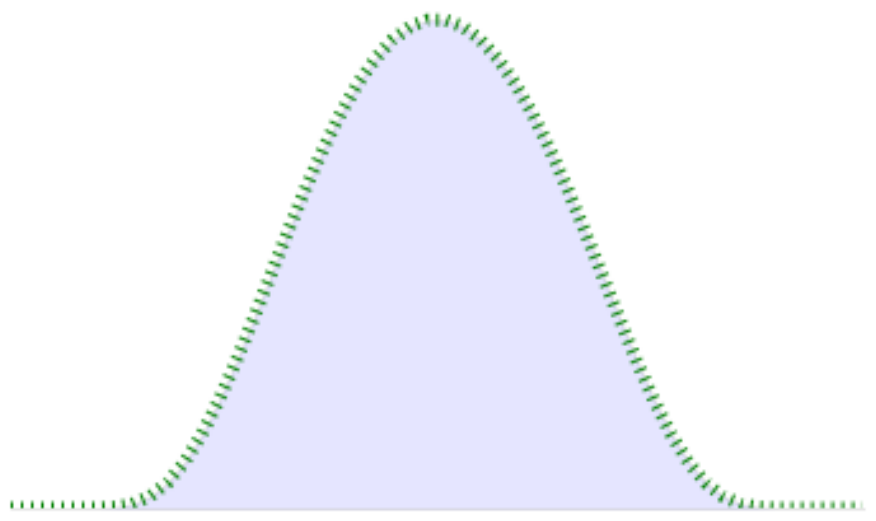
On *average* over a house
drawn from this distribution,
we want to make a good prediction.

Goal: predict future sale prices

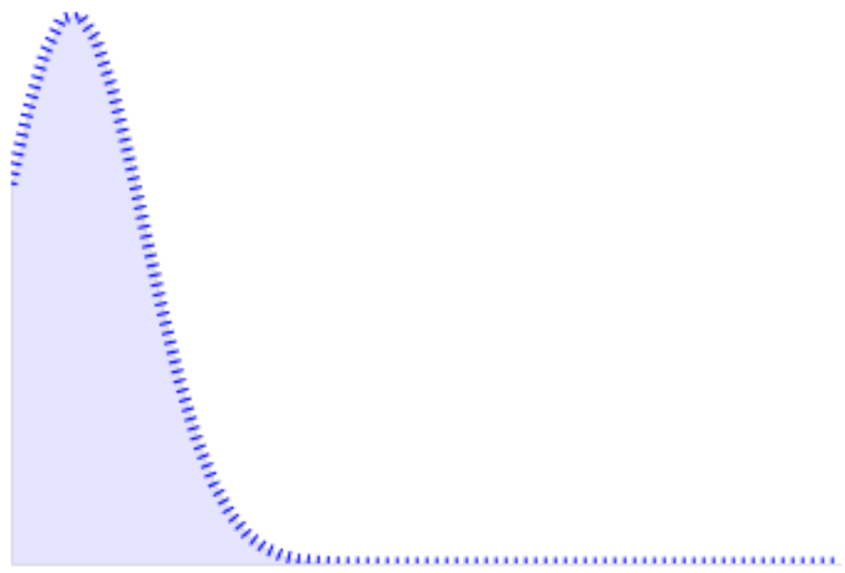
$$P_{XY}(X = x, Y = y)$$



$$P_{XY}(Y = y | X = x_0)$$



$$P_{XY}(Y = y | X = x_1)$$



Statistical Learning

$$P_{XY}(X = x, Y = y)$$

Goal: Predict Y given X

Find a function η that minimizes

$$\mathbb{E}_{XY}[(Y - \eta(X))^2]$$

Thus far, we've been using η which is a:
- Linear functions of X
- Degree p polynomials of X
- Linear "generalization" of X in p dimensions