BITS Pilani
Pilani Campus

BITS Pilani presentation

Tanmay Tulsidas Verlekar
CSIS

innovate achieve lead

BITS Pilani
Pilani Campus

Applied Machine Learning SE ZG568 / SS ZG568
Lecture No.5

The Bias/Variance Tradeoff

Bias

This part of the generalization error is due to wrong assumptions, such as assuming that the data is linear when it is actually quadratic. A high-bias model is most likely to underfit the training data.[10]

Variance

This part is due to the model's excessive sensitivity to small variations in the training data. A model with many degrees of freedom (such as a high-degree polynomial model) is likely to have high variance, and thus to overfit the training data.

Irreducible error

This part is due to the noisiness of the data itself. The only way to reduce this part of the error is to clean up the data (e.g., fix the data sources, such as broken sensors, or detect and remove outliers).

Regularized Linear Models

Equation 4-8. Ridge Regression cost function

$$J(\boldsymbol{\theta}) = \text{MSE}(\boldsymbol{\theta}) + \alpha\frac{1}{2}\Sigma_{i=1}^{n}\theta_i^{\,2}$$

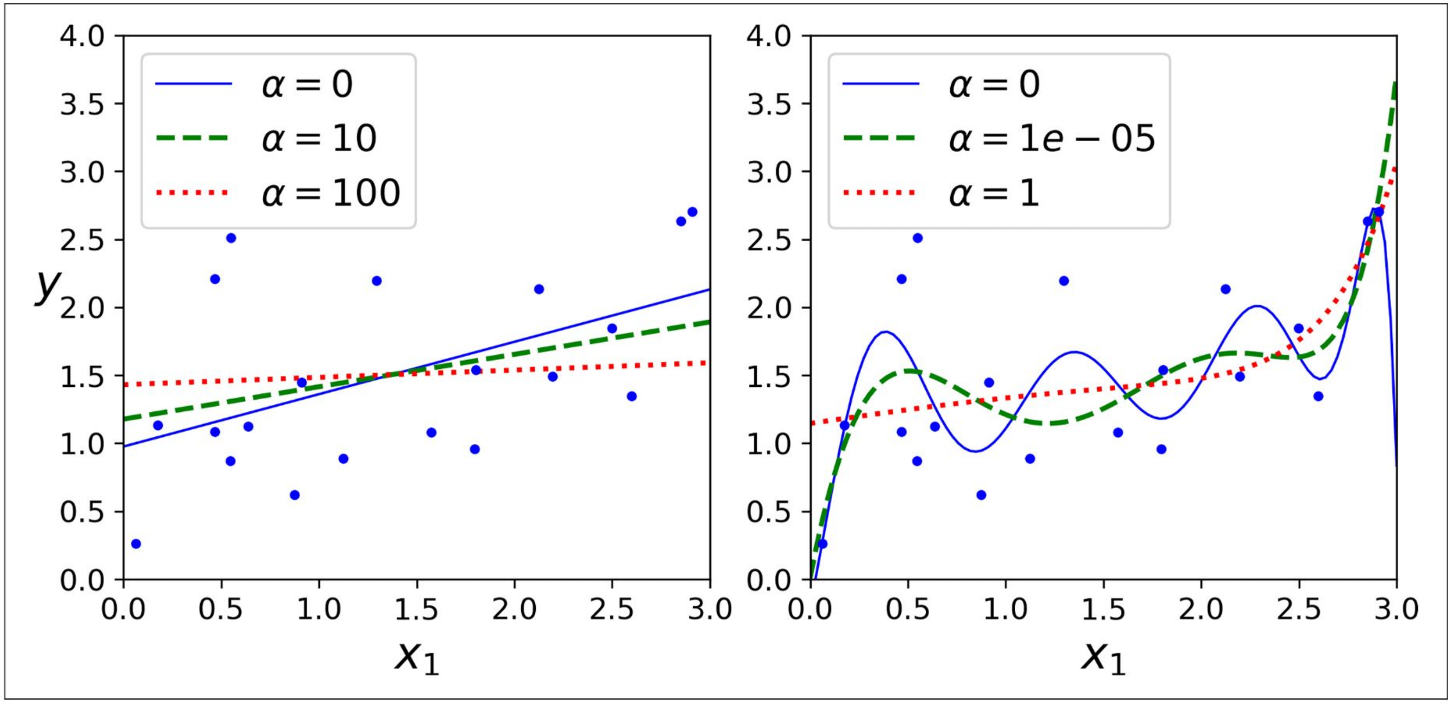

Figure 4-17. Ridge Regression

Lasso Regression

Equation 4-10. Lasso Regression cost function

$$J(\boldsymbol{\theta}) = \mathrm{MSE}(\boldsymbol{\theta}) + \alpha \sum_{i=1}^{n} \left|\theta_i\right|$$

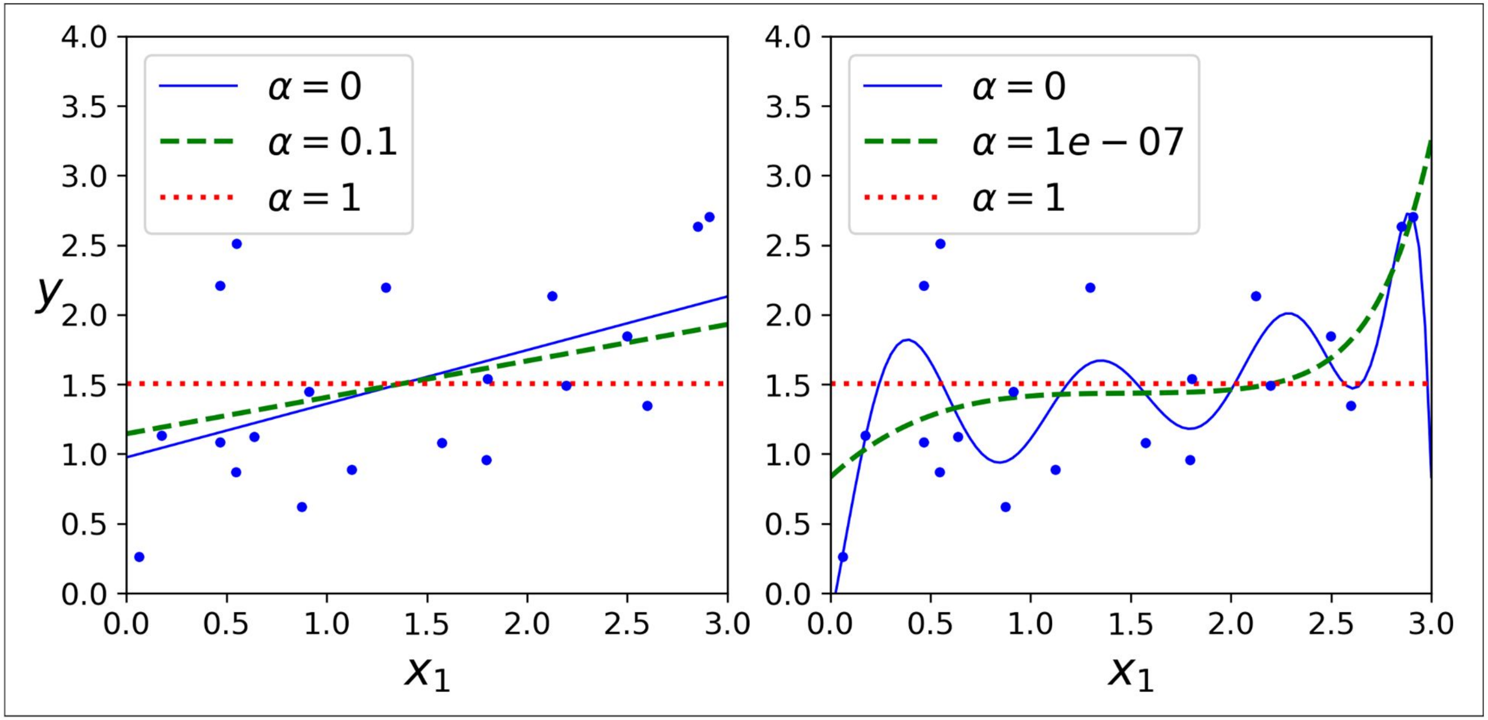

Figure 4-18. Lasso Regression

Elastic Net

Equation 4-12. Elastic Net cost function

$$J(\boldsymbol{\theta}) = \mathrm{MSE}(\boldsymbol{\theta}) + r\alpha\sum_{i=1}^{n}\left|\theta_i\right| + \frac{1-r}{2}\alpha\sum_{i=1}^{n}\theta_i^{2}$$

Early Stopping

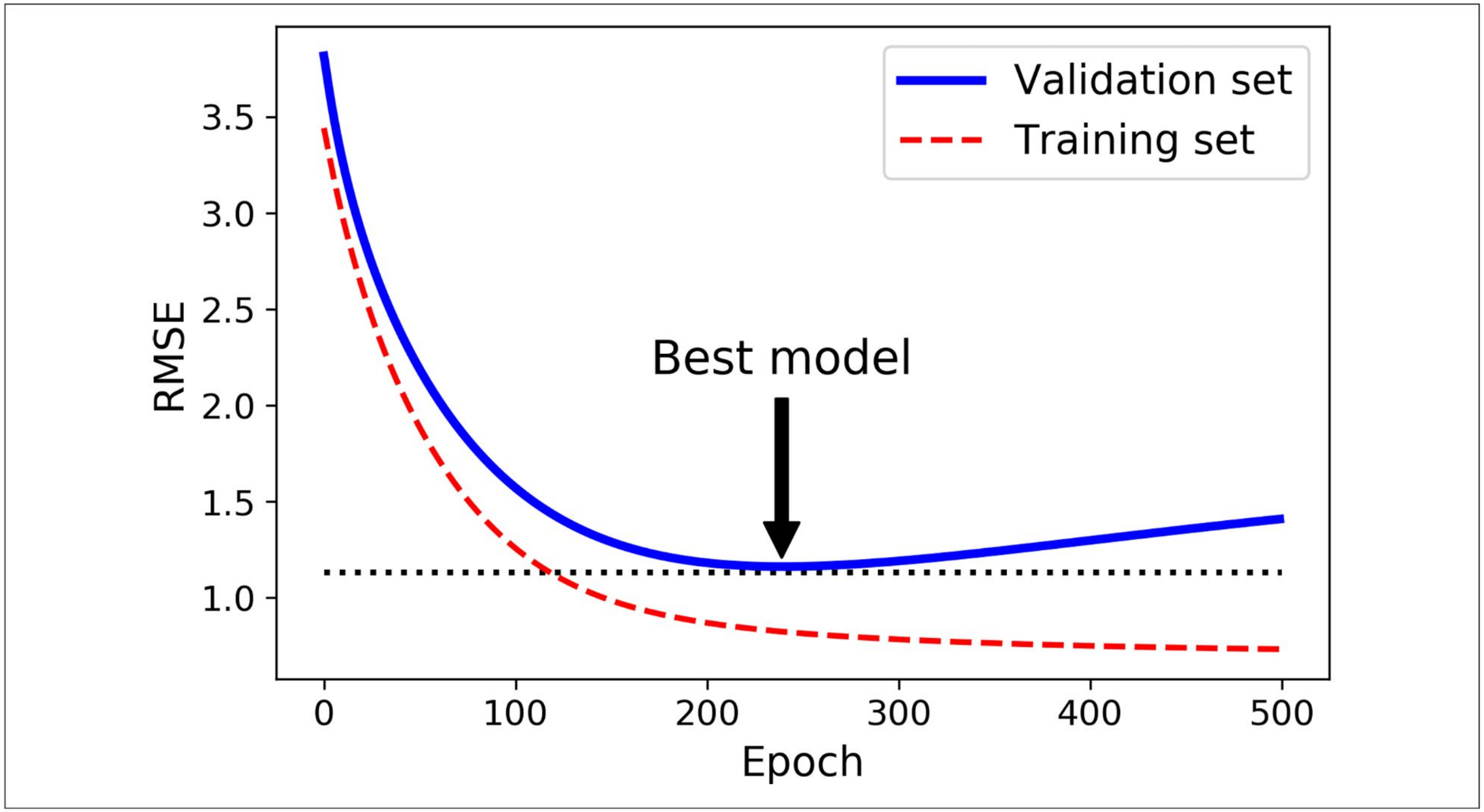

Figure 4-20. Early stopping regularization

Bayesian Classifiers

Relationship between the attribute set and the class variable is non-deterministic.

Although most people who eat healthily and exercise regularly have less chance of developing heart disease, they may still do so because of other factors such as heredity, excessive smoking, and alcohol abuse.

Bayes theorem

Bayes theorem (also known as the Bayes Rule or Bayes Law) is used to determine the conditional probability of event Y when event X has already occurred.

$$P(Y|X) = \frac{P(X|Y)P(Y)}{P(X)}.$$

Relationship between Joint and Conditional probability

Let X and Y be a pair of random variables. Their joint probability, $P(X = x, Y = y)$, refers to the probability that variable X will take on the value x and variable Y will take on the value y. A conditional probability is the probability that a random variable will take on a particular value given that the outcome for another random variable is known. For example, the conditional probability $P(Y = y|X = x)$ refers to the probability that the variable Y will take on the value y, given that the variable X is observed to have the value x. The joint and conditional probabilities for X and Y are related in the following way:

$$P(X, Y) = P(Y|X) \times P(X) = P(X|Y) \times P(Y). \tag{5.9}$$

Example

Consider a football game between two rival teams: Team 0 and Team 1. Suppose Team 0 wins 65% *of the time and Team 1 wins the remaining matches. Among the games won by Team 0, only* 30% *of them come from playing on Team 1's football field. On the other hand,* 75% *of the victories for Team 1 are obtained while playing at home. If Team 1 is to host the next match between the two teams, which team will most likely emerge as the winner?*

Probability Team 0 wins is $P(Y = 0) = 0.65$.
Probability Team 1 wins is $P(Y = 1) = 1 - P(Y = 0) = 0.35$.
Probability Team 1 hosted the match it won is $P(X = 1|Y = 1) = 0.75$.
Probability Team 1 hosted the match won by Team 0 is $P(X = 1|Y = 0) = 0.3$.

Solution

Our objective is to compute $P(Y = 1|X = 1)$, which is the conditional probability that Team 1 wins the next match it will be hosting, and compares it against $P(Y = 0|X = 1)$. Using the Bayes theorem, we obtain

$$
\begin{aligned}
P(Y = 1|X = 1) &= \frac{P(X = 1|Y = 1) \times P(Y = 1)}{P(X = 1)} \\
&= \frac{P(X = 1|Y = 1) \times P(Y = 1)}{P(X = 1, Y = 1) + P(X = 1, Y = 0)} \\
&= \frac{P(X = 1|Y = 1) \times P(Y = 1)}{P(X = 1|Y = 1)P(Y = 1) + P(X = 1|Y = 0)P(Y = 0)} \\
&= \frac{0.75 \times 0.35}{0.75 \times 0.35 + 0.3 \times 0.65} \\
&= 0.5738,
\end{aligned}
$$

$$P(A) = \sum_n P(A \cap B_n) \qquad P(A) = \sum_n P(A \mid B_n)P(B_n),$$

Prior and Posterior probabilities

Likelihood

Prior

Posterior

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Normalizing constant

$$P(B) = \sum_{Y} P(B|A)P(A)$$

A prior probability is the probability that an observation will fall into a group before you collect the data.

A posterior probability is the probability of assigning observations to groups given the data.

Example

	binary	categorical	continuous	class
Tid	**Home Owner**	**Marital Status**	**Annual Income**	**Defaulted Borrower**
1	Yes	Single	125K	No
2	No	Married	100K	No
3	No	Single	70K	No
4	Yes	Married	120K	No
5	No	Divorced	95K	Yes
6	No	Married	60K	No
7	Yes	Divorced	220K	No
8	No	Single	85K	Yes
9	No	Married	75K	No
10	No	Single	90K	Yes

Figure 5.9. Training set for predicting the loan default problem

Suppose we are given a test record with the following attribute set: $\mathbf{X} = (\texttt{Home Owner} = \texttt{No}, \texttt{Marital Status} = \texttt{Married}, \texttt{Annual Income} = \$120\text{K})$. To classify the record, we need to compute the posterior probabilities $P(\texttt{Yes}|\mathbf{X})$ and $P(\texttt{No}|\mathbf{X})$ based on information available in the training data. If $P(\texttt{Yes}|\mathbf{X}) > P(\texttt{No}|\mathbf{X})$, then the record is classified as Yes; otherwise, it is classified as No.

Bayes theorem is useful because it allows us to express the posterior probability in terms of the prior probability $P(Y)$, the **class-conditional** probability $P(\mathbf{X}|Y)$, and the evidence, $P(\mathbf{X})$:

$$P(Y|\mathbf{X}) = \frac{P(\mathbf{X}|Y) \times P(Y)}{P(\mathbf{X})}. \tag{5.11}$$

When comparing the posterior probabilities for different values of Y, the denominator term, $P(\mathbf{X})$, is always constant, and thus, can be ignored. The

Naıve Bayes Classifier

A naïve Bayes classifier estimates the class-conditional probability by assuming that the attributes are conditionally independent, given the class label y. The conditional independence assumption can be formally stated as follows:

$$P(\mathbf{X}|Y=y) = \prod_{i=1}^{d} P(X_i|Y=y), \tag{5.12}$$

where each attribute set $\mathbf{X} = \{X_1, X_2, \ldots, X_d\}$ consists of d attributes.

Naıve Bayes Classifier

To classify a test record, the naïve Bayes classifier computes the posterior probability for each class Y:

$$P(Y|\mathbf{X}) = \frac{P(Y)\prod_{i=1}^{d} P(X_i|Y)}{P(\mathbf{X})}. \tag{5.15}$$

Estimating Conditional Probabilities for Continuous Attributes

We can discretize each continuous attribute and then replace the continuous attribute value with its corresponding discrete interval.

We can assume a certain form of probability distribution for the continuous variable and estimate the parameters of the distribution using the training data.

$$P(X_i = x_i | Y = y_j) = \frac{1}{\sqrt{2\pi}\sigma_{ij}} \exp^{-\frac{(x_i - \mu_{ij})^2}{2\sigma_{ij}^2}}.$$

Solution

P(Home Owner=Yes|No) = 3/7
P(Home Owner=No|No) = 4/7
P(Home Owner=Yes|Yes) = 0
P(Home Owner=No|Yes) = 1
P(Marital Status=Single|No) = 2/7
P(Marital Status=Divorced|No) = 1/7
P(Marital Status=Married|No) = 4/7
P(Marital Status=Single|Yes) = 2/3
P(Marital Status=Divorced|Yes) = 1/3
P(Marital Status=Married|Yes) = 0

For Annual Income:
If class=No: sample mean=110
sample variance=2975
If class=Yes: sample mean=90
sample variance=25

$\texttt{No}$, $P(\texttt{Yes}) = 0.3$ and $P(\texttt{No}) = 0.7$.

$$\begin{aligned} P(\mathbf{X}|\texttt{Yes}) &= P(\texttt{Home Owner} = \texttt{No}|\texttt{Yes}) \times P(\texttt{Status} = \texttt{Married}|\texttt{Yes}) \\ &\quad \times P(\texttt{Annual Income} = \$120\text{K}|\texttt{Yes}) \\ &= 1 \times 0 \times 1.2 \times 10^{-9} = 0. \end{aligned}$$

$$\begin{aligned} P(\mathbf{X}|\texttt{No}) &= P(\texttt{Home Owner} = \texttt{No}|\texttt{No}) \times P(\texttt{Status} = \texttt{Married}|\texttt{No}) \\ &\quad \times P(\texttt{Annual Income} = \$120\text{K}|\texttt{No}) \\ &= 4/7 \times 4/7 \times 0.0072 = 0.0024. \end{aligned}$$