Prediction

Wenda Zhou

June 15, 2017

What is prediction?

Attempt to predict a random outcome with potentially unknown mechanism (may be viewed as random) from covariates.

Prediction in statistics

- One of fastest growing fields in statistics
- Driven by big data applications
- More used in non-scientific contexts

Prediction vs Estimation

Estimation

- Understand "true state of nature".
- Attempt to understand latent parameter

Prediction

- Attempt to guess another random quantity
- Prediction outcome is visible and observable

Prediction vs Estimation

Example

Clinical trial: estimation

Given a person with some characteristics, obtain the probability of success of the treatment on that person.

Prediction vs Estimation

Example

Clinical trial: estimation

Given a person with some characteristics, obtain the probability of success of the treatment on that person.

Clinical trial: prediction

For the next person that the doctor sees, predict whether the treatment will succeed.

Prediction vs Estimation

Example

Clinical trial: estimation

Given a person with some characteristics, obtain the probability of success of the treatment on that person.

Clinical trial: prediction

For the next person that the doctor sees, predict whether the treatment will succeed.

Two problems closely related.

What is a good prediction?

Define what a good prediction is.

Loss

Loss defines how bad a prediction was compared to the truth.

What is a good prediction?

Define what a good prediction is.

Loss

Loss defines how bad a prediction was compared to the truth.

Square loss

Suppose we predict $\hat{y}$, but the true value was y. The square loss is defined as:

$$L(y, \hat{y}) = (y - \hat{y})^2 \tag{1}$$

What is a good prediction?

Define what a good prediction is.

Loss

Loss defines how bad a prediction was compared to the truth.

Square loss

Suppose we predict $\hat{y}$, but the true value was y. The square loss is defined as:

$$L(y, \hat{y}) = (y - \hat{y})^2 \tag{1}$$

$0 - 1$ loss

For binary outcomes: 0 loss if we are right, 1 loss if we are wrong.

First example: prediction using linear models

Suppose that we have an ordinary linear model.
Then can predict a new outcome $\hat{y}$ with covariates x by putting:

$$\hat{y} = \hat{\alpha} + \hat{\beta}x \tag{2}$$

First example: prediction using linear models

Suppose that we have an ordinary linear model.
Then can predict a new outcome $\hat{y}$ with covariates x by putting:

$$\hat{y} = \hat{\alpha} + \hat{\beta}x \tag{2}$$

Unlike estimation, prediction is less worried about confounding.

Evaluating our prediction model

Easier to evaluate prediction models: can check whether we are right.
However, need care to separate the following quantities:

Training error

Loss incurred on training dataset.

Testing error

Loss incurred on new dataset.

Evaluating our prediction model

Warning

It is always true that:

$$\text{Training error} < \text{Testing error} \tag{3}$$

Evaluating our prediction model

Warning

It is always true that:

$$\text{Training error} < \text{Testing error} \quad (3)$$

In fact, training error is not a good measure of testing error.

Overfitting

Will often want to evaluate our model to select between multiple models.

Picking the model with lowest training error may not always be the best. May have overfitting.

Overfitting

Will often want to evaluate our model to select between multiple models.

Picking the model with lowest training error may not always be the best. May have overfitting.

Bias-variance and prediction

Although the bias-variance decomposition applies to estimation, prediction faces a very similar problem in overfitting vs. underfitting.

Classification

Take a binary (yes/no) decision – or discrete decision (which group is the subject in).
Very common case, so has specific vocabulary.

Classification

Sometime useful to separate the performance of classification in two parts:

Precision The precision p is the proportion of correct results among all predicted yes.

Recall The recall r is the proportion of correct results among all true yes.

Classification

Sometime useful to separate the performance of classification in two parts:

Precision The precision p is the proportion of correct results among all predicted yes.

Recall The recall r is the proportion of correct results among all true yes.

Sometimes useful to combine these measures into the $F1$ score:

$$F_1 = 2\frac{pr}{p+r} \tag{4}$$

Linear models for prediction

If have linear model:

$$y = \alpha + \beta x \tag{5}$$

Can generate predictions for new individual with covariates x_{new}:

$$\hat{y} = \hat{\alpha} + \hat{\beta} x_{\mathrm{new}} \tag{6}$$

Linear models for prediction

Prediction is less sensitive to confounding: $\hat{\beta}$ can be completely wrong and still helpful!

Linear models for prediction

Prediction is less sensitive to confounding: $\hat{\beta}$ can be completely wrong and still helpful!

Example: diamonds

Saw that diamonds with worse color rating were more expensive. Incorrect inference, but still helpful for prediction.

Uncertainty in prediction

Wish to quantify uncertainty of our prediction.

Prediction intervals

Interval likely to contain the true value of the outcome.

Uncertainty in prediction

Wish to quantify uncertainty of our prediction.

Prediction intervals

Interval likely to contain the true value of the outcome.

Unlike in estimation, even with infinite amounts of data, still uncertainty in prediction. Aggregate uncertainty in estimation and natural randomness of the outcome.