


Essential ML 102

☰ Tags	ML
➤ Class	
☑ Finished Yet?	☑
➤ Knowledge	 <u>The Seventh Sprint: Machine Learning</u>

-ข้อดีของ Text-based เทียบกับ Video-based:

1. เนื้อหาแบบ Text-based สามารถ Update + Maintain ได้ง่ายกว่า Video-based
2. โดยเฉลี่ยแล้ว นักเรียนสามารถเรียนจบได้ไวกว่า
3. Platform ที่มีชื่อเสียงในโลกออนไลน์ใช้ Text-based (มีวิดีโอแซมบ้างเล็กน้อย)
4. Text-based สามารถ Read & Copy นำไปใช้ได้ทันที

What is Algorithm?

-ML แต่ละตัวที่เราเรียน = Algorithm

-ให้ลองนึกภาพว่า Algorithm แต่ละประเภทคือโปเกมอน โปเกมอนแต่ละตัวมีจุดแข็ง จุดอ่อน แตกต่างกันไป Algorithm ก็เช่นกัน

- Algorithm แต่ละตัวจะมีวิธีการเรียนรู้ Pattern ในข้อมูลไม่เหมือนกัน (ศัพท์เทคนิคเรียกว่า Bias)
 - Algorithm บางตัวอธิบายผลได้ง่าย บางตัวเป็นเหมือนกล่องดำที่เราอธิบายอะไรแทบไม่ได้เลย แต่ตัวที่เราอธิบายแทบไม่ได้เลยมันกลับทายผลได้แม่นยำ
 - การเลือก ML Algorithm เป็นทักษะที่มาพร้อมกับประสบการณ์ ยิ่งเราเทรนโมเดลบ่อยเท่าไร เรา就越รู้ปัญหาที่นั้น ๆ ควรจะทดลองสร้างโมเดลแบบไหนบ้าง ไอเดียคือการทดลอง ลองผิดลอง ถูกซ้ำ ๆ ไปเรื่อย ๆ
- *Iterative process + Experimentation is key

Optional Reading - Ridge & Lasso (Explained by DataCamp)

pdf:

https://s3-us-west-2.amazonaws.com/secure.notion-static.com/97df787e-64a8-40d7-bfab-d09dabccf10a/Ridge_and_Lasso_Regression_-_DataCamp.pdf

- Ridge and Lasso are penalized regression methods (Introducing penalty factor to reduce overfitting)
- Lasso Equation:

Lasso Regression for Regularization

In this shrinkage technique, the coefficients determined in the linear model from equation 1.1. above are shrunk towards the central point as the mean by introducing a penalization factor called the alpha α (or sometimes lamda) values.

$$L_{lasso}(\hat{\beta}) = \sum_{i=1}^n (y_i - x_i' \hat{\beta})^2 + \lambda \sum_{j=1}^m |\hat{\beta}_j|.$$

Alpha (α) is the penalty term that denotes the amount of shrinkage (or constraint) that will be implemented in the equation. With alpha set to zero, you will find that this is the equivalent of the linear regression model from equation 1.2, and a larger value penalizes the optimization function. Therefore, lasso regression shrinks the coefficients and helps to reduce the model complexity and multi-collinearity.

Alpha (α) can be any real-valued number between zero and infinity; the larger the value, the more aggressive the penalization is.

-Ridge Equation:

Ridge Regression

Similar to the lasso regression, ridge regression puts a similar constraint on the coefficients by introducing a penalty factor. However, while lasso regression takes the magnitude of the coefficients, ridge regression takes the square.

<https://www.datacamp.com/tutorial/tutorial-lasso-ridge-regression>

4/18

10/25/22, 4:08 PM

Lasso and Ridge Regression Tutorial | DataCamp

$$L_{ridge}(\hat{\beta}) = \sum_{i=1}^n (y_i - x_i' \hat{\beta})^2 + \lambda \sum_{j=1}^m w_j \hat{\beta}_j^2.$$

Ridge regression is also referred to as **L2 Regularization**.

-Lasso = Least Absolute Shrinkage and Selection Operator [L1 regularization technique]

-Ridge = Lasso but put constraints on square instead of magnitude of coefficients [L2 regularization technique]

-In lasso, coefficients will be shrunk towards a mean of 0, less important features in a dataset are eliminated when penalized. The shrinkage of these coefficients based on

the alpha (or lambda) value provided leads to some form of automatic feature selection, as input variables are removed in an effective approach.

-Lasso: $\alpha * \sum(|\beta|)$

-Ridge: $\alpha * \sum(\beta^2)$

-Lasso can be used for model selection, but not Ridge

Conclusion

We have seen an implementation of ridge and lasso regression models and the theoretical and mathematical concepts behind these techniques. Some of the key takeaways from this tutorial include:

1. The cost function for both ridge and lasso regression are similar. However, ridge regression takes the square of the coefficients and lasso takes the magnitude.
 2. Lasso regression can be used for automatic feature selection, as the geometry of its constrained region allows coefficient values to inert to zero.
 3. An alpha value of zero in either ridge or lasso model will have results similar to the regression model.
 4. The larger the alpha value, the more aggressive the penalization.
-