Comparing Regression with Deep Learning and SVM

By: Peshal Pokhrel, Qudrat Ratul, Brett Sneed, and Maxwell Vestrand

Outline

Definitions

- What do Regression, SVM, and Deep Learning share?
- What is Regression?
- What is Deep Learning?
- What is Support Vector Machine (SVM)?

Recommendations

- When is SVM a better first choice than Regression?
- When is Regression a better first choice than SVM?
- When is Deep Learning a better first choice than Regression?
- When is Regression a better first choice than Deep Learning?

Conclusion

What do Regression, SVM, and Deep Learning share?

- Each is a Supervised Learning model: develop function from labelled training data which approximately maps input data to output data.

$$w_t \approx \text{argmin}_w \left\{ \sum_{(i,o)\in A} [E(o, f(i,w))] + R(w) \right\}$$

w	Model weights
w_t	Trained weights
A	Training data
i	Training data's input
o	Training data's output
$E(,)$	Error function for outputs
$f(,)$	Learning model
$R()$	Regularization function

What is Regression?

- Multiplies each input in ***i*** by its parameterization weight in ***w*** and then sums them
- Least squares as the error function

$$w_t \approx \text{argmin}_w \left\{ \sum_{(i,o)\in A} [E(o, f(i,w))] + R(w) \right\}$$

w	Parameterization weights
w_t	Optimized parameterization weights
A	Observed data
i	Observed input vector
o	Observed output
$E(,)$	Squared errror
$f(,)$	Weighted sum
$R()$	Regularization function

What is SVM?

- Quadratic Programming problem with constraints
- In typical real-world problems, classes will non-separable.
- In this case, have a trade-off between maximizing margin ***R()*** and minimizing training error ***E(,)***.
- Hinge loss:

$$H(z) = \max(0, 1 - z)$$

$$w_t \approx \text{argmin}_w \left\{ \sum_{(i,o)\in A} [E(o, f(i, w))] + R(w) \right\}$$

w	Hyperplane class margin
w_t	Trained hyperplane class margin
A	Labelled training data
i	Training data's input vector
o	One-hot encoding of class
$E(,)$	Hinge loss function
$f(,)$	Classification by side of margin
$R()$	Margin-size regularization function

What is Deep Learning?

- Each vector of connection weights ***w*** maps to a connection between layers
- Deep learning architecture including activation functions are encapsulated in ***f(,)***
- To be true “Deep” learning, need at least say 10 layers

$$w_t \approx \text{argmin}_w \left\{ \sum_{(i,o)\in A} [E(o, f(i, w))] + R(w) \right\}$$

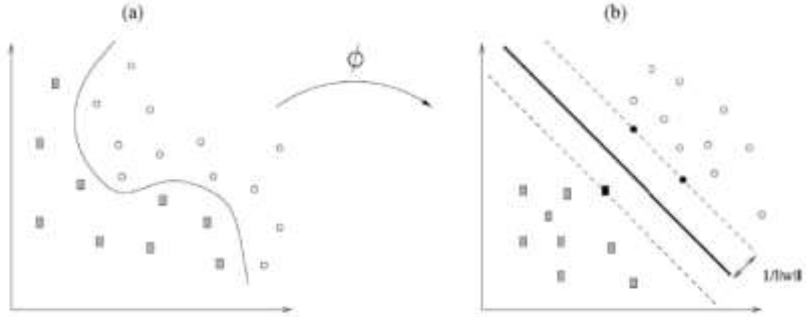
w	Model layer connection weights
w_t	Trained layer connection weights
A	Labelled training data
i	Training data’s input vector
o	One-hot encoding of class
$E(,)$	Sum of squared errors
$f(,)$	Evaluation of deep learning network
$R()$	Regularization function

When is SVM a better first choice than Logistic Regression?

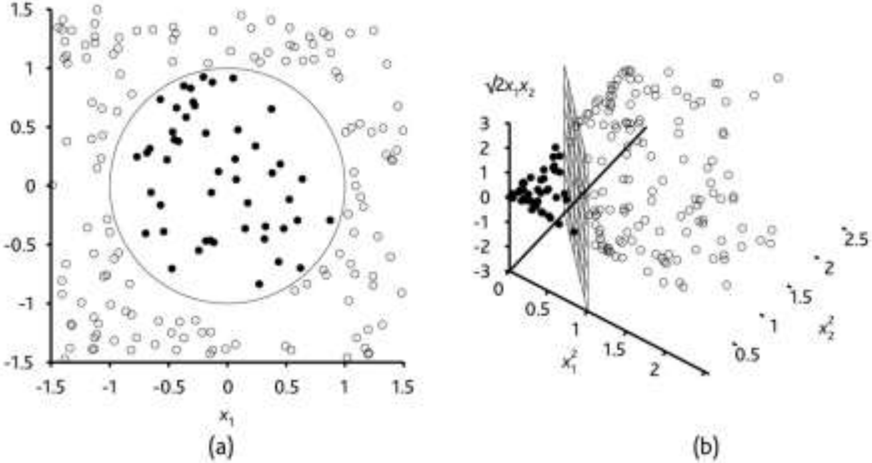
SVM is better when

- High **dimension** data
- It is effective in cases where number of dimensions is greater than the number of samples.
- Works really well with **clear margin of separation**
- Uses a subset of training points in the decision function, so its **memory efficient**

SVM vs LR



Transformation using Kernel Function



Transformation to higher dimension

Source: https://bmcmedinformdecismak.biomedcentral.com/articles/10.1186/1472-6947-8-56

https://www.researchgate.net/publication/260773845_Comparison_between_SVM_and_Logistic_Regression_Which_One_is_Better_to_Discriminate

Example-1

mortality prediction for haematological malignancies patients.

[https://bmcmedinformdecismak.biomedcentral.com/articles/10.1186/1472-6947-8-56]

Total Patients: 352

Training Data : 252

Testing Data: 100

Input variable: 17

First Model: 12 input variable

Second Model: 17 input variable

Input variable	Training	Validation
age, yrs	55 (± 18)	58 (± 15)
% high-grade malignancy	61	54
% active disease of relapse	34	39
% allogeneic bone marrow transplant./stem cell transplant.	13	10
days of hospitalisation before ICU admission, median (IQR)	4 (16)	6 (16)
% bacterial infection	44	43
pulse (/min)	123 (± 28)	118 (± 33)
mean blood pressure (MAP), mmHg	73 (± 27)	69 (± 22)
respiration frequency (/min)	32 (± 10)	33 (± 13)
Pa02/Fi02 (p/f)	198 (± 130)	194 (± 126)
platelets (1000/mm^3)	125 (± 700)	90 (± 114)
urea<24 h (g/l)	0.86 (± 59)	0.82 (± 55)
creatinine<24 h (mg/dl)	1.6 (± 1.08)	1.7 (± 1.7)

Example : Comparison

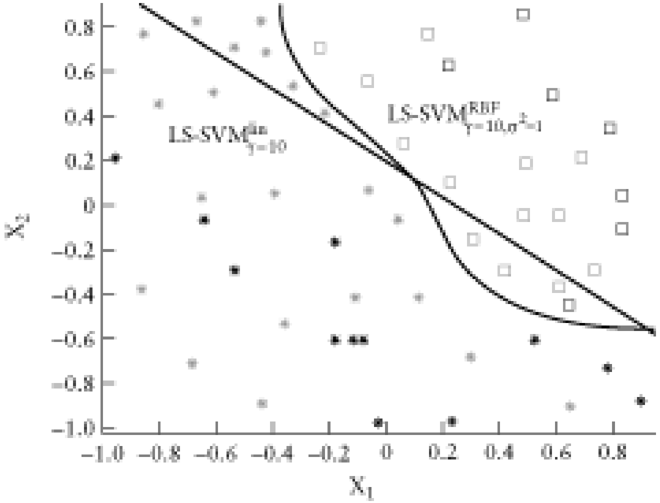
[https://bmcmedinformdecismak.biomedcentral.com/articles/10.1186/1472-6947-8-56]

	MLR -1	SVM -1	MLR - 2	SVM - 2
Accuracy	0.730	0.680	0.740	0.680
positive predictive value	0.755	0.740	0.780	0.739
Negative predictive value	0.702	0.630	0.700	0.630

Source: [https://bmcmedinformdecismak.biomedcentral.com/articles/10.1186/1472-6947-8-56]

SVM

Radial Basis Function Vs Least Squares Support Vector Machine



Source: [http://onlinelibrary.wiley.com/doi/10.1002/uog.2791/full]

When is Logistic Regression a better than SVM

SVM vs LR

Regression is good at Working large amount of data efficiently.

Regression use all data; SVM use marginal data

Why Logistic Regression

- To avoid overfitting
  - Large amount of noise or class overlap
  - Large number of features for a relatively small number of examples
- Logistic regression and linear kernel SVM produce similar results
  - Similar mathematical justifications
  - Logistic regression considers all points, SVM considers only the nearest points
  - Logistic regression may be more accurate when training set has unbalanced distribution of classes in the training data

When is Deep Learning a better first choice than Regression?

Some Applications where Deep Learning Out performs

Colorization of Black and white Images

Adding sounds to the Silent movies

Automatic Machine Translation

Deep Dreaming

Colorization of Black and white Images

Using Convolutional NN

Adding sounds to the Silent movies

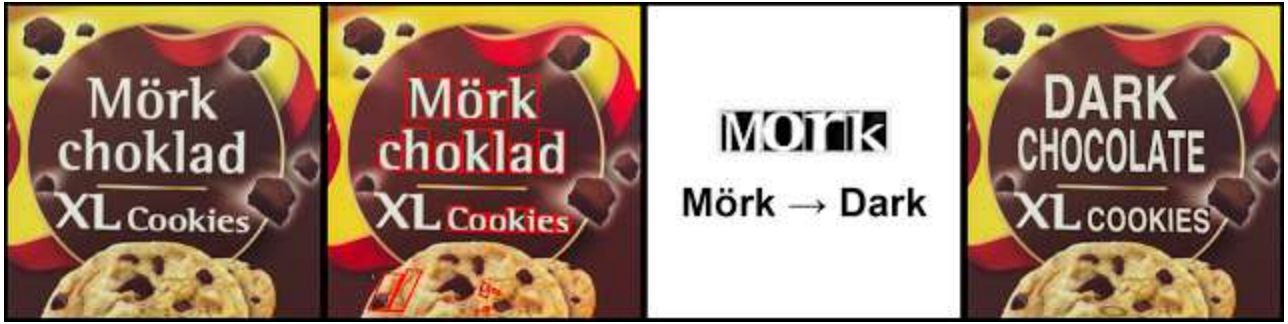
A set of video is trained to the system hitting in different surface.

The system was able to detect whether the surface is hard, soft liquid etc.

You Tube: Visually Indicated sound (MITCSAIL)

Automatic Machine Translation

Instant Visual Translation



Source: Google research blog

Deep Dreaming

Source: Deep dreaming generator

When is Regression a better first choice than Deep Learning?

If the output is numerical, try Regression.

- If trying to predict a numerical value with unbounded range (like Housing price), try Regression
- Predicting a numerical value is different from assigning to a discrete class

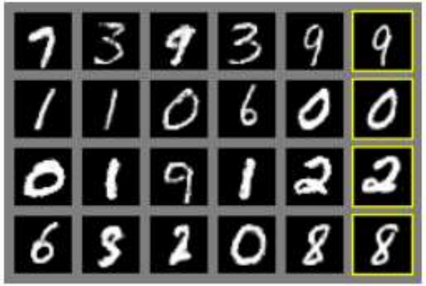
Input				Output
Square Footage	Number of Bedrooms	Number of Bathrooms	Size of garage in cars	Housing Price
2,500	3	2.5	2	$250,000
1,500	1	1.5	1	$120,000
4,500	3	4	3	$1,000,000
950	1	1	0	$80,000

If the data set is small, try Regression.

“As of 2016, a rough rule of thumb is that a supervised deep learning algorithm will generally achieve acceptable performance with around **5,000 labeled examples per category**, and will match or exceed human performance when trained with a dataset containing at least 10 million labeled examples.”

- Goodfellow et al. 2016, **Deep Learning**

- Regression models can develop good predictions from only 2 to 10 data points depending on the complexity of the regression model.



Generative Adversarial Nets,
Goodfellow et al. 2014

If you’re only working on one CPU, try Regression.

“Traditionally, neural networks were trained using the CPU of a single machine. Today, this approach is generally considered insufficient. **We now mostly use GPU computing or the CPUs of many machines networked together.** Before moving to these expensive setups, researchers worked hard to demonstrate that CPUs could not manage the high computational workload required by neural networks.”

- Goodfellow et. al. 2016, **Deep Learning**

- The global minimum of the error function for Linear Regression can be found by solving a small system of equations in a few seconds on a single CPU.

Conclusion

- Regression, Deep Learning, and SVM are each supervised learning algorithms with their own strengths and weaknesses.
- Common advice: start with the simplest model that solves the problem.
- Then, advance the complexity of the model over time to meet needs/goals.