Inference of subsurface properties by machine learning

Huy Le (huyle@stanford.edu)

CS 229 FALL 2017

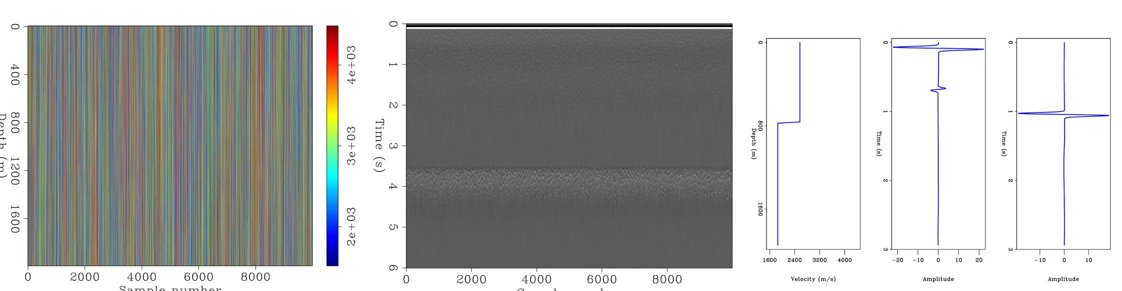
Motivation

Inference of the Earth's properties from seismic data is conventionally performed by solving an optimization problem that minimizes the difference between observed data and simulated data using gradient descent methods. In this project, I propose to use machine learning to predict subsurface properties, particularly rocks' acoustic velocity and interfaces' depth, from recorded seismic data.

Data and labels

10000 unique one-dimensional velocity profiles are generated randomly and their corresponding seismic data are simulated. 100 profiles are set aside as a test set. Two sensors are used to record seismic waves: one at the surface and one at the bottom of the velocity profile. Sampling rate is 4 milliseconds.

The seismic waveforms are my raw data. I later processed this data to obtain two arrival times, one for reflection and one for transmission, and one reflection coefficient. I also labeled these data in two ways: first, using only three parameters (two velocities and one depth) and second, 160 discrete velocity values at every 12.5 m.



(a) (b) (c)

Figure 1: (a) 10000 unique randomly generated one-dimensional velocity profiles. (b) Recorded seismic data: direct and reflection waves from 0-3 seconds and transmitted waves from 3-6 seconds. (c) One sample velocity profile and its data.

Methods

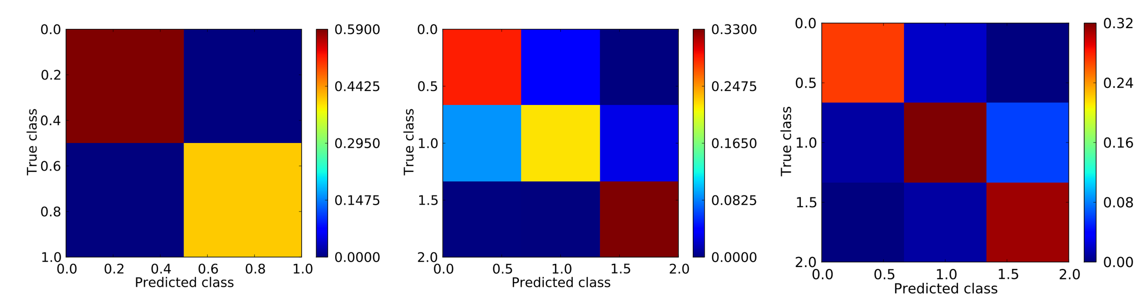
I use a dense neural network with one hidden layer of 100 neurons. Given that there is a relationship between the my labels (the velocity profiles) and my data (the seismic traces) through the wave equation, neural networks seem to be a good method to approximate this relationship.

First, I formulate the problem as a classification one with two depth classes and three velocity classes. For this I use sigmoid activation function for the hidden layer and softmax function for the output layer with cross-entropy loss function. Second, I consider the problem as a regression. For this case, I use relu activation function for the hidden layer and none for the output layer with least-square loss function. For both cases, I use mini-batch gradient descent to minimize the loss functions.

As a classification problem

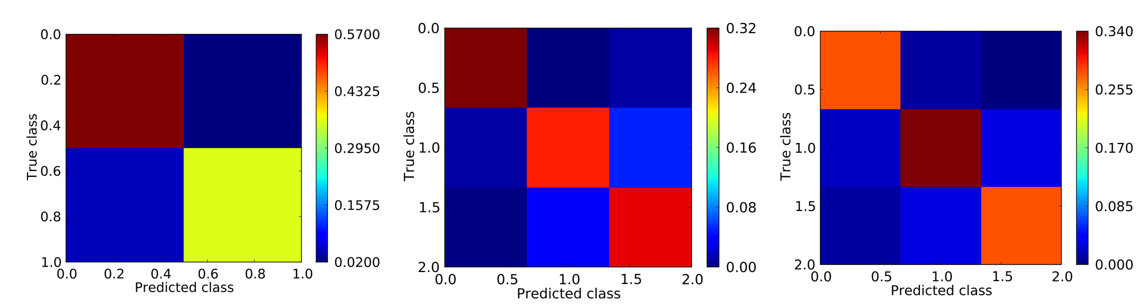
Depth classes	0-1 km	1-2 km	
Velocity classes	1.5-2.5 km/s	2.5-3.5 km/s	3.5-4.5 km/s

Table 1: Two classes of depth and three classes of velocity. Since my parameters consist of one depth and two velocities, total number of classes is $2 \times 3 \times 3 = 18$.



(a) (b) (c)

Figure 2: Confusion matrices for depth (a), v_1 (b), and v_2 (c) when processed data (i.e. arrival times and reflection coefficient) are used as features. These matrices are diagonally dominant, saying that there are more correct classifications than misses. Note that the confusion matrix for depth is actually diagonal, meaning depth is classified perfectly.

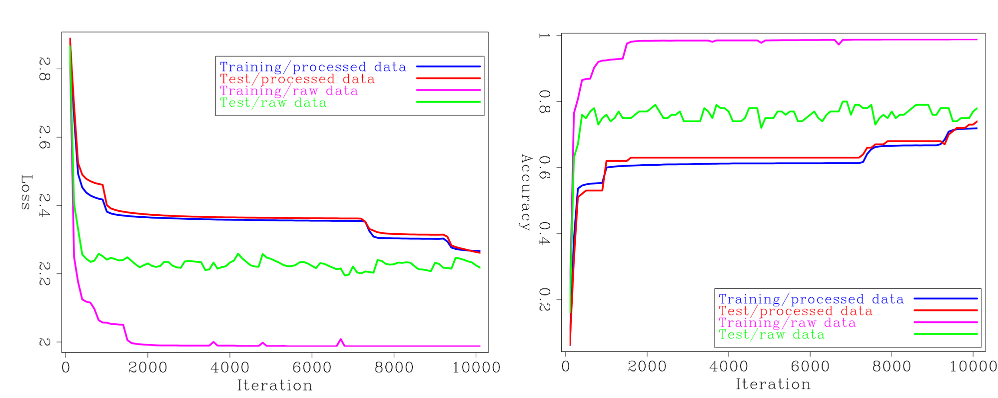


(a) (b) (c)

Figure 3: Confusion matrices for depth (a), v_1 (b), and v_2 (c) when raw waveform data are used as features.

Data type	Total miss	Depth miss	v_1 miss	v_2 miss
Processed	26	0	16	10
Raw	22	7	11	10

Table 2: Summary of classification results. Size of test set is 100. When arrival times and reflection coefficient are used as features, reflector's depth is classified perfectly and mis-classifications occur either in velocity above or below. When the raw waveforms are used, there is slightly less mis-classifications, but they occur in all parameters.



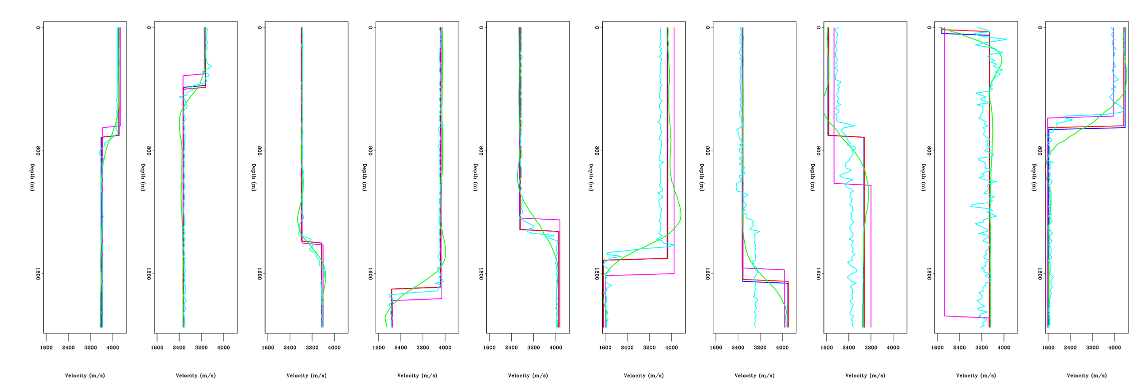
(a) (b)

Figure 4: Loss function (a) and accuracy (b) of training data and test data for classification experiments. One interesting observation is that when feature is the raw waveform, training loss reduces and accuracy improves compared to processed data as features, but test loss and accuracy don't differ much for both types of data. Another observation is the loss function curves and accuracy curves seem to mimic each other.

As a regression problem

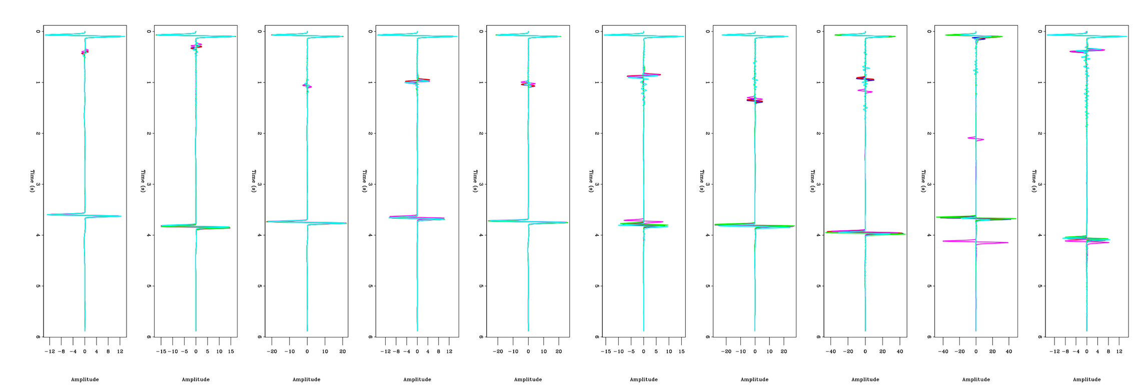
Data type	Label type	Average velocity error	Color
processed	three parameters	23.54 m/s	red
processed	discrete	75.04 m/s	green
raw	three parameters	140.15 m/s	magenta
raw	discrete	104.94 m/s	cyan

Table 3: Summary of regression results. Prediction with processed data (i.e. two arrival times and one reflection coefficient) and three-parameter label gives the best result while using raw data to predict three-parameter label is worst.



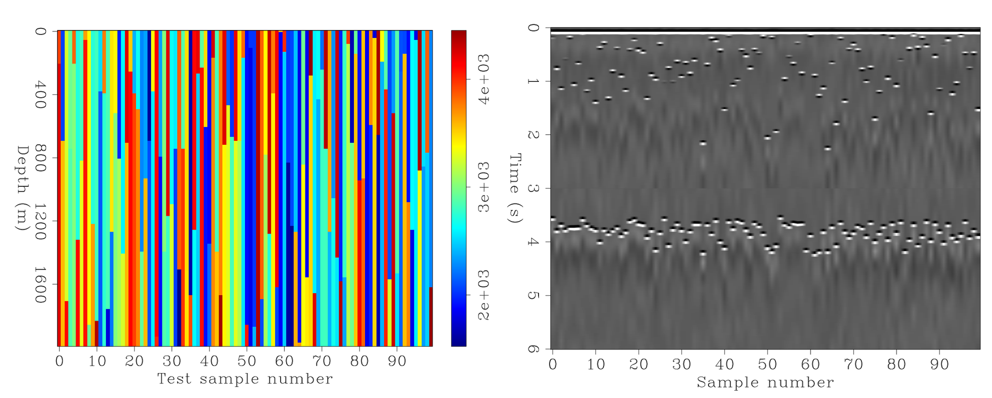
(a) (b)

Figure 5: Five relatively good (a) and unsuccessful (b) predictions. The true profiles are in blue. See Table 3 for other color codes. Interestingly, when processed data and discrete velocity label are used, the network predicts smooth velocity profiles (green curves).



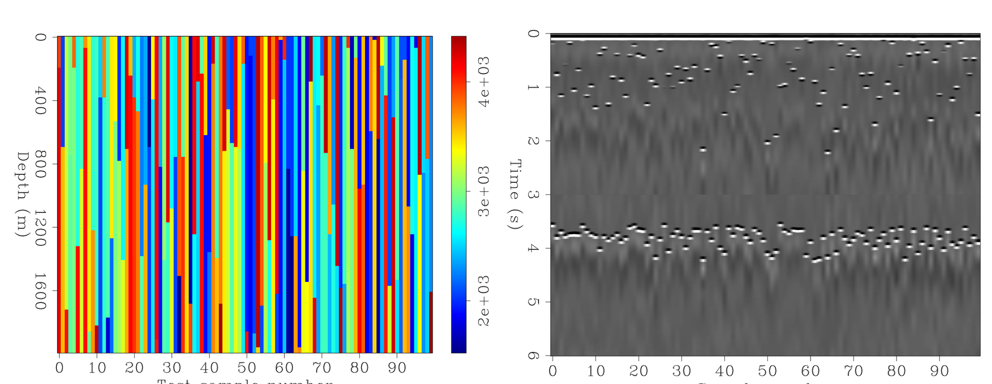
(a) (b)

Figure 6: Simulated data from good (a) and failed (b) predictions. Good predictions match both reflection and transmission events while unsuccessful ones mostly fail to predict reflections and sometimes both events.



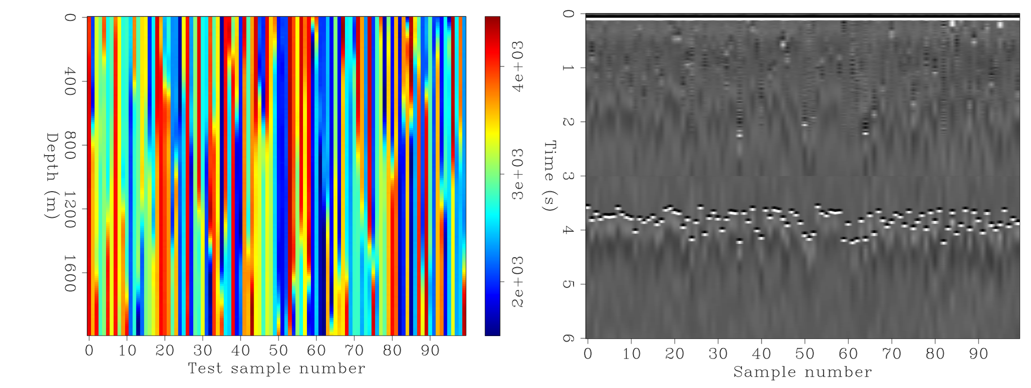
(a) (b)

Figure 7: (a) Test profiles and (b) simulated data.



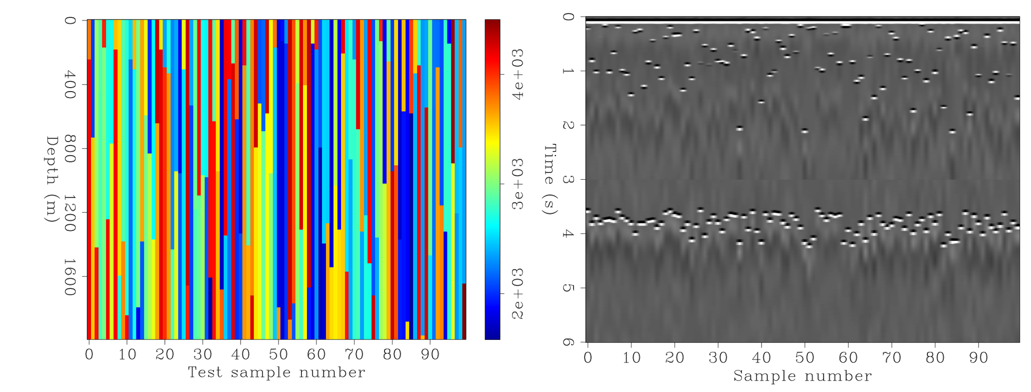
(a) (b)

Figure 8: (a) Predicted profiles and (b) simulated data when using processed data and three-parameter label. Compare with Figure 7.



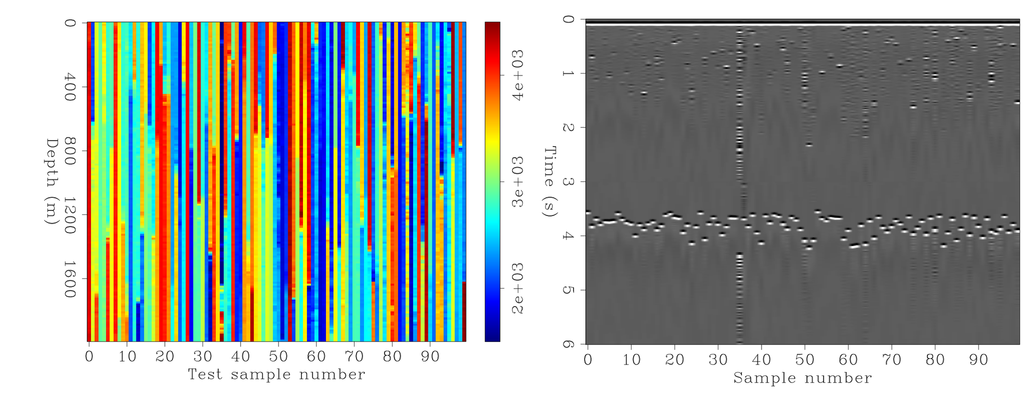
(a) (b)

Figure 9: (a) Predicted profiles and (b) simulated data when using processed data and discrete-velocity label. Compare with Figure 7.



(a) (b)

Figure 10: (a) Predicted profiles and (b) simulated data when using raw data and three-parameter label. Compare with Figure 7.



(a) (b)

Figure 11: (a) Predicted profiles and (b) simulated data when using raw data and discrete-velocity label. Compare with Figure 7.

Conclusions

1. For the classification problem:
   - both types of data have similar performance on the test set, roughly 75% accuracy,
   - when processed data is used as features, the network classifies reflector's depth perfectly and only misses either one of two velocities,
   - when raw waveform is used, mis-classifications occur in all parameters.
2. For the regression problem:
   - prediction with processed data and three-parameter label gives the best result while using raw waveforms to predict three-parameter label is worst,
   - when processed data and discrete velocity label are used, the network predicts smooth velocity profiles.

References

[1] Gunter Roth and Albert Tarantola, 1994, Neural networks and inversion of seismic data, Journal of Geophysical Research, Vol. 99, No. B4, 6753-6768.