

**NANYANG  
TECHNOLOGICAL  
UNIVERSITY**  

---

**SINGAPORE**

**EE6222: MACHINE VISION  
Assignment 1 Option 1**

**RVFL Report**

**Author:**

Tan Chuan Xin (U1821755B)

**Date of submission:**

7 November 2021

# Datasets

**You can use around 8-10 datasets. If your computer cannot handle large datasets with more than 20000 samples, you can exclude them. You should also very exclude small datasets. You can also have some 2 class and more than 2 class problems.**

Ten datasets were selected from the UCI Data Python folder. The machine that the code was running on does not have a dedicated GPU, hence computing resources are constrained, and the maximum dataset size is <5000 samples to achieve a feasible runtime.

The following are the datasets that have been selected, along with their respective sizes

Dataset	Dataset size (rows, columns)
abalone	(4177, 10)
bank	(4521, 18)
car	(1728, 8)
dermatology	(366, 36)
echocardiogram	(131, 12)
glass	(214, 11)
seeds	(210, 9)
teaching	(151, 7)
titanic	(2201, 5)
wine	(178, 15)

Very small datasets that are <100 rows have been excluded from the analysis.

## RVFL Code

**Using this RVFL as the base model, you are asked to investigate various issues as listed in your lecture slides. You can refer to your lecture slides for further details on RVFL.**

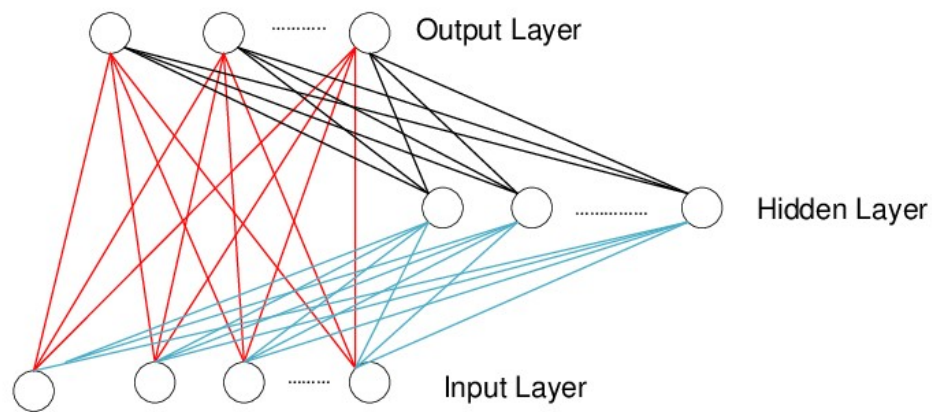
Building upon the base code provided, modifications have been made to adapt the code to answer the three questions posed in the assignment. The two rounds of notifications about errors in the base code have been noted, accounted for, and resolved accordingly.

For each of the following six RVFL models, a hyperparameter sweep has been done according to the base code to optimise the number of neurons in the hidden layer (N), scaling factor (S), and the regularization parameter (C) for the respective models. This optimizes the model's performance.

Subsequently, the direct links, activation function and regularization mode are chosen based on the question being answered.

# Question 1

Effect of direct links from the input layer to the output layer (i.e. with and without)



The direct links from the input layer to the output layer in a RVFL are represented by the red links. A traditional neural network, or multilayer perceptron model does not have direct connections between the input layer and the output layer. However, hidden layer weights need to be learnt and trained using back propagation. This makes RVFL a faster network to train.

In RVFL, the hidden layer weights (blue lines) are randomized and fixed, and we need to learn the weights of the red and black links. Question 1 investigates if the direct links actually help the model learn better. The following settings have been implemented, along with the results obtained. The results are in the form of **(cross validation accuracy  $\pm$  variance)**.

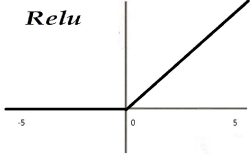
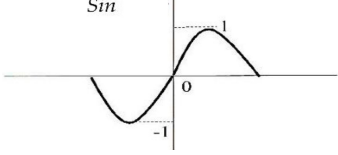
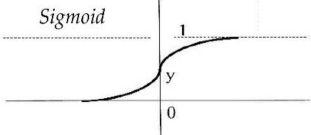
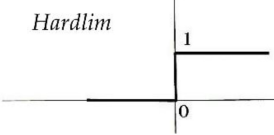
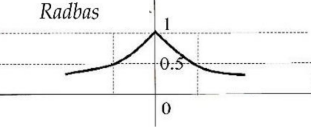
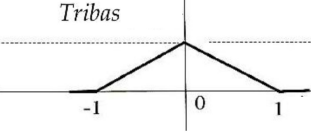
	Without direct link	With direct link	Higher accuracy with direct link?	Lower variance with direct link?
Direct link	False	True		
Activation function	radbas	radbas		
Regularization	Ridge regression	Ridge regression		
<b>Dataset</b>				
abalone	0.660 $\pm$ 1.74E-04	0.662 $\pm$ 1.17E-04	Yes	Yes
bank	0.899 $\pm$ 1.29E-05	0.900 $\pm$ 1.04E-05	Yes	Yes
car	0.953 $\pm$ 1.34E-06	0.953 $\pm$ 1.21E-05	Same	Yes
dermatology	0.967 $\pm$ 6.04E-05	0.981 $\pm$ 8.30E-05	Yes	No
echocardiogram	0.818 $\pm$ 2.75E-03	0.856 $\pm$ 1.72E-04	Yes	Yes
glass	0.642 $\pm$ 4.81E-03	0.670 $\pm$ 1.51E-03	Yes	Yes
seeds	0.952 $\pm$ 8.32E-04	0.947 $\pm$ 6.24E-04	No	Yes
teaching	0.612 $\pm$ 2.55E-03	0.572 $\pm$ 4.98E-03	No	No
titanic	0.786 $\pm$ 1.63E-05	0.790 $\pm$ 3.03E-06	Yes	Yes
wine	0.983 $\pm$ 3.55E-04	0.983 $\pm$ 3.55E-04	Same	Same
AVERAGE	0.827 $\pm$ 1.16E-03	0.831 $\pm$ 7.87E-04	Yes	Yes

From the table above, it is evident that direct links tend to help models achieve higher accuracies and lower variance. Therefore, we not only achieve better predictions, but we also achieve more consistent predictions. There is only one case where the use of direct links harmed the model in both accuracy and variance, and that is with the **teaching** dataset.

Nonetheless, it is safe to conclude that direct links between the input and output layer are likely to have beneficial effects for the models.

## Question 2

**Performance comparisons of 2 activation functions: one from “relu, sigmoid, radbas, sine” and one from “hardlim, tribas”**

Activation function	Illustration	Activation function	Illustration
Relu $y = \max(0, s)$		Sine $y = \sin(s)$	
Sigmoid $y = \frac{1}{1 + e^{-s}}$		Hardlim $y = \frac{(\text{sign}(s) + 1)}{2}$	
Radbas $y = e^{-s^2}$		Tribas $y = \max(1 -  s , 0)$	

Activation functions are applied at neurons in the RVFL in order to achieve non-linearity. These activation functions form the basis of the universal approximator property of RVFLs. The default activation function set in the base code is radbas, and therefore we shall select radbas, and hardlim and investigate the effect on model performance when changing the activation function.

Direct links for this question will be set to true, in accordance to the default settings for the RVFL model in the base code. Also, it follows from question 1 that direct links are generally beneficial for the models.

	radbas	hardlim	Higher accuracy with hardlim?	Lower variance with hardlim?
Direct link Activation function Regularization	True radbas Ridge regression	True hardlim Ridge regression		
Dataset				
abalone	0.662 ± 1.17E-04	0.647 ± 4.51E-05	Yes	Yes
bank	0.900 ± 1.04E-05	0.894 ± 1.54E-05	No	No
car	0.953 ± 1.21E-05	0.862 ± 3.88E-05	No	No
dermatology	0.981 ± 8.30E-05	0.984 ± 3.02E-05	Yes	Yes
echocardiogram	0.856 ± 1.72E-04	0.871 ± 1.72E-04	Yes	Same
glass	0.670 ± 1.51E-03	0.698 ± 3.03E-03	Yes	No
seeds	0.947 ± 6.24E-04	0.947 ± 6.24E-04	Same	Same
teaching	0.572 ± 4.98E-03	0.559 ± 1.30E-04	No	Yes
titanic	0.790 ± 3.03E-06	0.789 ± 1.41E-05	No	No
wine	0.983 ± 3.55E-04	0.994 ± 9.68E-05	Yes	Yes
AVERAGE	0.831 ± 7.87E-04	0.816 ± 4.20E-04	No	Yes

Generally, the hardlim activation function gives slightly lower accuracies versus radbas. However, it provides slightly lower variance overall than radbas. It should be noted that both accuracies and variances show very little difference between hardlim and radbas.

Nonetheless, we can observe that though the effects of hardlim on accuracy and variance are opposite, the drop in accuracy when using hardlim is more significant than the decrease in variance, given the difference in the order of magnitude at which they operate on.

Therefore, we can conclude that radbas is the better activation function, at least when comparing against these ten datasets.

## Question 3

**Performance of Moore-Penrose pseudoinverse and ridge regression (or regularized least square solutions) for the computation of the output weights.**

Ridge regression	Moore-Penrose pseudoinverse
$\beta = D(D^T D + \lambda I)^{-1} T$	$A^+ = (A^T A)^{-1} A^T$

For ridge regression, we are trying to solve the problem of L2 regularization, where we adjust the regularization parameter  $\lambda$  to impose penalties on the model. Moore-Penrose pseudoinverse is a method that allows for computation of a best fit solution for the least-squares problem, even if there is no true solution. It also provides a regularization effect for the synaptic outputs.

We similarly set direct links to true following the result of question 1, and activation function to radbas, following the result of question 2.

Direct link Activation function Regularization	Ridge regression True radbas Ridge regression	Moore-Penrose True radbas Moore-Penrose	Higher accuracy with Moore- Penrose?	Lower variance with Moore- Penrose?
Dataset				
abalone	$0.662 \pm 1.17\text{E-}04$	$0.657 \pm 6.45\text{E-}05$	No	Yes
bank	$0.900 \pm 1.04\text{E-}05$	$0.895 \pm 9.59\text{E-}06$	No	Yes
car	$0.953 \pm 1.21\text{E-}05$	$0.953 \pm 1.21\text{E-}05$	Same	Same
dermatology	$0.981 \pm 8.30\text{E-}05$	$0.975 \pm 8.30\text{E-}05$	No	Same
echocardiogram	$0.856 \pm 1.72\text{E-}04$	$0.871 \pm 1.72\text{E-}04$	Yes	Same
glass	$0.670 \pm 1.51\text{E-}03$	$0.443 \pm 3.65\text{E-}03$	No	No
seeds	$0.947 \pm 6.24\text{E-}04$	$0.962 \pm 1.85\text{E-}04$	Yes	Yes
teaching	$0.572 \pm 4.98\text{E-}03$	$0.618 \pm 2.25\text{E-}03$	Yes	Yes
titanic	$0.790 \pm 3.03\text{E-}06$	$0.790 \pm 3.03\text{E-}06$	Same	Same
wine	$0.983 \pm 3.55\text{E-}04$	$0.955 \pm 7.75\text{E-}04$	No	No
AVERAGE	$0.831 \pm 7.87\text{E-}04$	$0.812 \pm 7.20\text{E-}04$	No	Yes

Generally, the Moore-Penrose pseudoinverse gives lower accuracies versus ridge regression. However, it provides slightly lower variance overall than ridge regression.

We can observe that though the effects of Moore-Penrose pseudoinverse on accuracy and variance are opposite, the drop in accuracy when using Moore-Penrose pseudoinverse is more significant than the decrease in variance, given the difference in the order of magnitude at which they operate on.

Therefore, we can conclude that ridge regression is the better activation function, at least when comparing against these ten datasets.

## Conclusion

Following the experiments conducted, we can conclude the following:

- Q1 – direct links are better than no direct links
- Q2 – radbas is generally better than hardlim
- Q3 – ridge regression is generally better than Moore-Penrose pseudoinverse

It should be noted that these results are only valid for the ten datasets presented in this report. The generalizability of radbas/ridge regression is debatable because these experiments do not provide conclusive evidence for their superiority over other options. Direct links however are evidently very effective, much more than not having direct links.