Automatic cancer development prediction based on classification of mass lesions in mammograms

Roman Roman, Homero
Stanford University
homero@stanford.edu

2016-11-18

1 Abstract

Beyond simple classification of breast mammograms as normal, benign, and cancerous this project also aims to use the learned classification models to create prediction images of the development of normal regions to cancerous ones. These prediction images are done trough reinforcement learning whereby the normal mammogram is tweaked enough to be classified as benign and then from benign to malignant. This can also be done backwards to trace back the development of cancerous regions and try to reconstruct how the area used to look like when it was healthy.

2 Introduction

"About 40,450 women in the U.S. are expected to die in 2016 from breast cancer, though death rates have been decreasing since 1989 ... the result of treatment advances, earlier detection through screening, and increased awareness" (National Breast Cancer Foundation) As such, developing effective detection methods both for diagnosis and prevention is a vital part of combating breast cancer. In this project, diagnosis is done through multiclass classification of mamammographs into normal, benign, and cancerous while the prevention characterization is done by the automatic prediction of cancer development through reinforcement learning.

3 Data

The data for this project comes from the Digital Database for Screening Mammography (DDSM), a resource maintained by the University of South Florida. The database contains over 2,000 studies, where each study is two images of each breast and associated patient information. One image is from the Cranial-Caudal (CC) view (the view of the breast from above) and the other from the medio-lateral view (MLO) view (a side-angle view). The database is found at:

http://marathon.csee.usf.edu/Mammography/Database.html

For convenience, the benign and malignant lesion images are taken from the following Dropbox repository

https://www.dropbox.com/s/oz9a40unwitvy27/DDSM.zip

where the images have been cropped into areas of interest of about 600 x 600 pixels

4 Classification Hypothesis

Since we want to classify into three labels for a mammogram (normal, benign, malignant), a multi-class classification model seems appropriate for the situation. The state-of-the-art approach nowadays is to use a neural network, but in this project simple k-Nearest Neighbors clustering and Softmax Regression will also be used for comparison.

5 Implementation

5.1 k-Nearest Neighbors

For the k-Nearest Neighbors classification, the training images are stored as pixel matrices and to classify a test image we find the closest k train images by Euclidean distance. Then we take the average classification of these neighbors to classify the test image.

For example if we have three pixel-size train images A:{0,0,0} and B:{0,1,0} and C:{ 1,1,1} and labels 1 (normal), 2(benign), 3(cancerous). Then for k=1 and test image T:{1,0,1} we take Euclidean distances AT = $sqrt(2)$, BT = $sqrt(3)$ CT = $sqrt(1)$ and we would classify T as 2 (benign)

While pretty straightforward, this classification approach does not fully capture the image properties and thus is expected to have high error rates (though it is not expected to over-fit). As such, as an algorithm that performs better than random this can be treated as the baseline.

5.2 Softmax regression

5.2.1 Data pre-processing

For Softmax, one can use the raw image pixels as features but it is also possible to extract the following features:
pixel by pixel intensity,
local standard deviation,
local average

One can also run k-means multiple times with random initialization and Silhouette validation of consistency to find the most stable number of clusters and place the pixels into these clusters.

5.2.2 Softmax Model

After doing data pre-processing one can fit a Softmax regression model for k=3 (Since we have 3 labels) where the labels $y^(i) \in \{1, 2, ..., k\}$. More specifically, we do gradient descent on the following cost function[1] including a weight decay term to ensure the cost function $J(\theta)$ is convex for any hyper-parameter $\lambda > 0$

$$J(\theta) = -\frac{1}{m}[\sum_{i=1}^{m}\sum_{j=1}^{k} 1\{y^{(i)} = j\} log \frac{e^{\theta_j^T x^{(i)}}}{\sum_{l=1}^{k} e^{\theta_l^T x^{(i)}}}] + \frac{\lambda}{2}\sum_{i=1}^{k}\sum_{j=0}^{n}\theta_{ij}^2$$

and taking the gradient gives:

$$\nabla_{\theta_j} J(\theta) = -\frac{1}{m}\sum_{i=1}^{m}[x^{(i)}(1\{y^{(i)} = j\} - \frac{e^{\theta_j^T x^{(i)}}}{\sum_{l=1}^{k} e^{\theta_l^T x^{(i)}}})] + \lambda\theta_j$$

then plugging into the stochastic gradient descent rule we get:

$$\theta_j := \theta_j - \alpha\nabla_{\theta_j} J(\theta)$$

for each value $j = 1, ..., k$
where we set $\alpha := 0.001$ $\lambda := \frac{1}{m}$ $m :=$ Number of Train Images and $k := 3$
Finally we use this model to assign a label of 1 (normal), 2(benign), or 3(cancerous).

5.3 Still in development: Neural Network Classification

Since much research has already been done on Neural Net classification, one can employ already developed libraries such as PyBrain using the multinomial logistic regression gradient to classify the images. One point of interest with this approach to explore in the future is the ranges of the hyper-parameters which give best results.

5.4 Experimental Results

Thus far, the preliminary version of the algorithm only works for k-Nearest Neighbors. After leave-one-out cross validation, the confusion matrix results are as follows for 500 normal, 500 benign, and 500 cancerous mammograms:

	PREDICTED NORMAL	PREDICTED BENIGN	PREDICTED CANCEROUS
ACTUAL NORMAL	365	95	60
ACTUAL BENIGN	66	300	90
ACTUAL CANCEROUS	45	100	379

Accuracy = 1044/1500 = 0.696

5.5 Future Work: Generation of cancer development prediction images by reinforcement learning

After the model has successfully learned how to distinguish between normal, benign, and malignant, we take normal mammograms and proceed to tweak them enough to be classified as benign and then malignant. The states are how the entire image looks, the actions are to increase or decrease a single pixel intensity and we assign transition probabilities by modeling the center of clusters as behaving like the centers of normal distributions such that points closer to cluster centers will be visited more often.

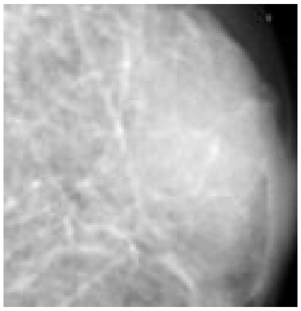

Figure 1: INITIAL NORMAL

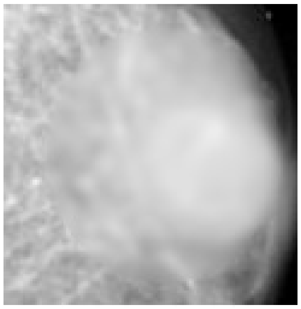

Figure 2: GENERATED BENIGN

References

[1] National Breast Cancer Foundation $http : //www.breastcancer.org/symptoms/understand_bc/statistics$

[2] $http : //cs229.stanford.edu/proj2015/278_poster.pdf$

[3] Unsupervised Feature Learning and Deep Learning. "Softmax Regression" $http : //ufldl.stanford.edu/wiki/index.php/$ $Softmax_Regression$