week7_code

Created: 18/11/24 09:04 **Updated:** 18/11/24 23:52

partial codes in week_7 code assignment.

> written by *VincentX3*, Nov.24.18

gaussianKernel.m

```
dist=sum((x1-x2).^2);
sim=exp(-dist/(2*sigma^2));
```

dataset3Params.m

```
C_list=[0.01, 0.03, 0.1, 0.3, 1, 3, 10, 30];
sigma_list=[0.01, 0.03, 0.1, 0.3, 1, 3, 10, 30];
lc=length(C_list);
lsigma=lc;
errors=ones(length(C_list)*lsigma,1);

%debug
format_debug='indice: i=%4d,j=%4d,subscript=%4d\n';

for i=1:lc
    for j=1:lsigma
        fprintf(format_debug,i,j,(i-1)*lc+j);

        model= svmTrain(X, y, C_list(i), @(x1, x2) gaussianKernel(x1, x2, sigma_list(j)));
        predictions = svmPredict(model, Xval);

        %use Determine inequality to measure.
        %A ~= B returns a logical array with elements set to logical 1 (true)
        %where arrays A and B are not equal; otherwise, the element is logical 0(false).
        %matlab's subcript begin with 1, so remember to 'i-1'
        errors((i-1)*lc+j)=mean(double(predictions ~= yval))
    end
end

error_min=min(errors);
subscript=find(errors==error_min);
C=C_list(floor(subscript/lc));
sigma=sigma_list(mod(subscript,lsigma));
```

- use Determine inequality :A ~= B returns a logical array
- matlab's subcript begin with 1, so remember to 'i-1'

```
%better version
%from @github/benoitvallon
%sincerly thanks.
bestPrediction = 1000;
for C_i = [0.01, 0.03, 0.1, 0.3, 1, 3, 10, 30]
    for sigma_i = [0.01, 0.03, 0.1, 0.3, 1, 3, 10, 30]
    model= svmTrain(X, y, C_i, @(x1, x2) gaussianKernel(x1, x2, sigma_i));
    predictions = svmPredict(model, Xval);
    prediction = mean(double(predictions ~= yval));
    if prediction < bestPrediction;
        bestPrediction = prediction;
        C = C_i;
        sigma = sigma_i;
        end
    end
end
```

Spam Filter

1. --normalize-- the email
2. extract features
3. compare with vocabulary
4. training

---

most of the SVM library:no need to add this extra feature x0 = 1 yourself.