NAME: ID:

Reminders (do not ask what variables stand for):

$$\mathcal{N}\left(x|\mu,\sigma^2\right) = \frac{1}{(2\pi\sigma^2)^{1/2}} \exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\}$$

$$\mathcal{N}(\mathbf{x}|\boldsymbol{\mu},\boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{D/2}} \frac{1}{|\boldsymbol{\Sigma}|^{1/2}} \exp\left\{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right\}$$

Sigmoid(x) $= \frac{1}{1+e^{-x}}$

Threshold(x)= 1 if $x \geq 0$
0 otherwise

RELU(x) = x if $x \geq 0$
0 otherwise

Binary-cross-entropy loss: $-(y.\log \hat{y} + (1-y).\log(1-\hat{y}))$

Entropy of X:

$$H[X] = -\sum_{x_i} p(X = x_i).log2(p(X = x_i))$$

Cross-Entropy between p and q:

$$H(p,q) = -\sum_{x_i} p(X = x_i).log2(q(X = x_i))$$

Cross-entropy loss: $-\mathbf{y}.\log(\hat{\mathbf{y}})$

NAME: ID:

1) 12pt – Basic Math for ML

a) 3pts – You have trained a linear regression model $\hat{f}(x) = a + bx$ that minimizes the squared distance between the target value y and predicted value $\hat{f}(x)$, for a sample x.

What is the **mean square error** of the estimates $\hat{f}$(x) over the given test set T ={(x_i,y_i)}, i=1..M.

<u>Give as a formula, no partial for errors/missing details.</u>

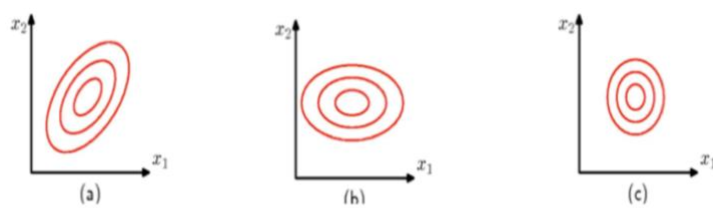
MSE = ...

b) 3pts – Apply gradient descent to **minimize** the function f(x,y) = x^3 + 2xy + 4 starting at the point p_0 = (x y)= $[1\ 2]^T$. Use a learning rate of 0.1 and just go one step.

Gradient =

p_1 =

c) 3pts – Match given covariance matrices for the following 3 Normal distributions which are shown with their equal density contours. Use arrows to indicate the matchings.



$\Sigma = \begin{pmatrix} 4 & 0 \\ 0 & 10 \end{pmatrix}$ $\Sigma = \begin{pmatrix} 9 & 0 \\ 0 & 4 \end{pmatrix}$ $\Sigma = \begin{pmatrix} 16 & 13 \\ 13 & 20 \end{pmatrix}$

d) **3pt –** Assume you are given the following conditional probability distributions for left handedness with respect to gender.

P(LeftHanded |Male) = 0.3. P(LeftHanded |Female) = 0.2

Compute the probability of picking/encountering a LeftHanded person from the whole population. <u>You must show your work.</u> Assume:

P(Male) = 0.4 P(Women)=0.6

Answer: ....

NAME: ID:

5) 10pts – Bayesian Approaches

a) **6pt -** Consider a naive Bayes classifier trained on the dataset given below, where **all features are binary**.

Sample	Fever	Runny Nose	Throat Pain	Body Ache	Disease
1	High	No	Yes	Yes	Flu
2	High	No	No	Yes	Flu
3	Normal	No	Yes	Yes	Flu
4	Normal	No	Yes	No	Flu
5	High	No	Yes	No	Cold
6	Normal	Yes	Yes	No	Cold
7	Normal	Yes	No	Yes	Cold
8	Normal	Yes	Yes	No	Cold

A new patient comes with the following symptoms:

x = [Fever=Normal, RunnyNose=No, ThroatPain=No, BodyAche=Yes)

a) 3pts - **Give the probabilities needed for estimating** P(Flu | X). I.e. write the required probabilities as P(A|B) etc. and not as numbers.

P(Flu | x) ∝ P(...............................) x P(..........................) x. P(...............................) x

P(...............................) x P(...............................)

Answer expected: P(Flu | x) ∝ P(Fever=Normal|Flu)xP(RN=No|Flu)xP(TP=No|Flu)xP(BA=Yes|Flu)xP(Flu)

b) **3pts -** Now plugin the numbers calculated from the table (<u>do not use smoothing</u>) as fractions, but <u>leave out the multiplication</u>

P(Flu | x) ∝ ....

Answer expected: P(Flu | x) ∝ 2/4 x 4/4 x …..

b) 2pts - Lets assume that you have found that P(Flu |x) = 0.2/P(x) and P(Cold | x) = 0.3/P(x) in the above problem.

What would be the actual normalized value of P(Flu | x)?

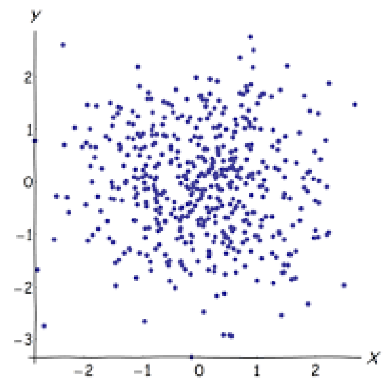
Answer: ……………………………………………….

c) 2pt –Select all correct answers (each wrong cancel one correct):

- ○ P (Cold | x) + P (Flu| x) = 1
- ○ P (x |Cold) + P (x | Flu) = 1
- ○ Naive Bayes uses a simplifying assumption that attributes are conditionally independent given the class label
- ○ Naive Bayes uses a simplifying assumption that classes are conditionally independent given the attributes

NAME: ID:

6) 8pts – Multivariate Normal Distribution



a) 4pts – Assume you are given the above two-dimensional data, in the form of a scatter plot.

- **1pt –** What can you say about its value (positive?, negative?, 0?, 1?, -1?, …)? *Be as precise as possible.*

- **1pts –** What is the range of correlation**, in general**? I.e. what numbers can it take? **Answer**: …………………

- **1pts –** What is the range of variance, **in general?** I.e. what numbers can it take? **Answer**: …………………

…

NAME: ID:

8) 8pts –Classifier Combination

a) 5pts - Circle as True or False. 1pts each correct; -1pts for each incorrect.

- **T / F -** In bagging, **the training set for each base model** may not contain some of the samples in the original training set and may contain a single sample from the original training multiple times.
- **T / F -** Bagging is **not very suitable** for **stable** learning algorithms such as k-NN.
- **T / F -** Ensemble **averaging** is **guaranteed** to have higher accuracy than each of the base classifiers.
- **T / F -** In **random forest,** we select to use a subset of the features for **each tree**.
- **T / F -** **The main advantage** of combining classifiers is reducing **variance.**

b) 3pts – Considering a **majority vote** ensemble on a two class problem, assume that each of the **5 base classifiers** has a probability of error ***p =0.2*** on a given input and that their errors are independent.

What can you say about the probability of error for the ensemble?

- Probability that the ensemble makes error < Probability of exactly 3 base classifiers make errors
- Probability that the ensemble makes error > Probability of exactly 3 base classifiers make errors
- Probability that the ensemble makes error = Probability of exactly 3 base classifiers make errors
- Cannot say

Hint: You can answer this question without using any formulas, with just reasoning

NAME: ID:

10) 15pts – General ML

a) 5pts – A bank loses about 1 million TL a year on fraud cases and wants to detect credit card fraud (fraud or not) using machine learning techniques. Credit card fraud occurs in about 5% of all the bank's credit card transactions.

i. 2pts - What is the **majority classifier's accuracy** in this classification problem?

Answer: ……………………

ii. 3pts - What is the **random classifier's accuracy** in this classification problem?

Answer: ……………………