KTH VETENSKAP OCH KONST

KTH Electrical Engineering

Solutions to Exam in Pattern Recognition EN2202

Date: Monday, Oct 18, 2010, 08:00 – 13:00

Place: V35, L1.

Allowed: Beta (or corresponding), calculator with empty memory. No notes!

Grades: A: 31p; B: 27p; C: 23p; D: 20p; E: 17; of max 25p + 10p project bonus.

Language: Swedish or English.

Results: Friday Nov 5, 2010.

Review: At KTH-S3/ STEX, Osquldas v. 10.

Good Luck!

Please do the **Course Evaluation**! See the course web page.

1 In a given pattern-classification application there are two source categories, here called $S = 1$ and $S = 2$. These two source types are known to occur with equal probabilities. The classifier input is a feature vector $\boldsymbol{X} = (X_1, X_2)^T$ with two elements that are *non-negative* and statistically *independent* of each other. Depending on the source category $S = i$, each feature vector element X_k, with $k = 1$ or $k = 2$ has an *exponential* conditional distribution, with probability density functions of the form

$$f_{X_k|S}(x_k|i) = \begin{cases} \lambda_{ik} e^{-\lambda_{ik} x_k}, & 0 \leq x_k \\ 0, & x_k < 0 \end{cases}$$

The distribution parameters are exactly known:

$$S = 1 : \quad \begin{cases} \lambda_{11} = 1 \\ \lambda_{12} = 2 \end{cases} \qquad S = 2 : \quad \begin{cases} \lambda_{21} = 2 \\ \lambda_{22} = 1 \end{cases}$$

(a) Design an optimal classifier that can guess the source category with minimum error probability, and simplify the classifier to show that it is possible to make optimal decisions using a single *linear* discriminant function of the type $g(x_1, x_2) = ax_1 + bx_2 + c$, with a threshold mechanism. (3p)

Solution: As both source alternatives are equally probable, we use the *Maximum Likelihood* decision rule. As the two feature elements are independent, the feature-vector density is just the product of the density functions for the feature elements. We can use a single discriminant function

$$\begin{aligned} g(\boldsymbol{x}) &= \ln f_{\boldsymbol{X}|S}(\boldsymbol{x}|1) - \ln f_{\boldsymbol{X}|S}(\boldsymbol{x}|2) = \\ &= \ln \lambda_{11} - \lambda_{11} x_1 + \ln \lambda_{12} - \lambda_{12} x_2 - \ln \lambda_{21} + \lambda_{21} x_1 - \ln \lambda_{22} + \lambda_{22} x_2 = \\ &= \ln \frac{\lambda_{11}\lambda_{12}}{\lambda_{21}\lambda_{22}} + (\lambda_{21} - \lambda_{11})\, x_1 + (\lambda_{22} - \lambda_{12})\, x_2 = x_1 - x_2 \end{aligned}$$

Thus, the classifier should use the decision rule

$$d(x_1, x_2) = \begin{cases} 1, & x_1 > x_2 \\ 2, & \text{otherwise} \end{cases}$$

(b) What is the the conditional probability that the source category was $S = 1$, given an observed feature vector $\boldsymbol{x} = (1, 2)^T$? (1p)

Solution: Using Bayes rule, we have

$$\begin{aligned} P(S = 1|\boldsymbol{X} = \boldsymbol{x}) &= \frac{f_{\boldsymbol{X}|S}(\boldsymbol{x}|1)}{f_{\boldsymbol{X}|S}(\boldsymbol{x}|1) + f_{\boldsymbol{X}|S}(\boldsymbol{x}|2)} = \\ &= \frac{\lambda_{11} e^{-\lambda_{11} x_1} \lambda_{12} e^{-\lambda_{12} x_2}}{\lambda_{11} e^{-\lambda_{11} x_1} \lambda_{12} e^{-\lambda_{12} x_2} + \lambda_{21} e^{-\lambda_{21} x_1} \lambda_{22} e^{-\lambda_{22} x_2}} = \\ &= \frac{e^{-x_1 - 2x_2}}{e^{-x_1 - 2x_2} + e^{-2x_1 - x_2}} = \frac{e^{-5}}{e^{-5} + e^{-4}} = \frac{1}{1 + e} \end{aligned}$$

(c) What is the probability of correct decisions, using the optimal classifier? (1p)

Solution: As both source probabilities are equal, the probability of correct decision is

$$P_c = P(d(\boldsymbol{X}) = 1|S = 1)P(S = 1) + P(d(\boldsymbol{X}) = 2|S = 2)P(S = 2) = P(d(\boldsymbol{X}) = 1|S = 1)$$

The decision region R_1 for $d = 1$ is the half quadrant between the line $x_2 = 0$ and the diagonal line $x_1 = x_2$. The conditional probablity of observing a feature vector in this region, given $S = 1$, is

$$\begin{aligned}
P(X \in R_1|S = 1) &= \int_0^\infty \int_0^{x_1} f_{X_1|S}(x_1|1) f_{X_2|S}(x_2|1) dx_2 dx_1 \\
&= \int_0^\infty f_{X_1|S}(x_1|1) \left[\int_0^{x_1} f_{X_2|S}(x_2|1) dx_2 \right] dx_1 = \\
&= \int_0^\infty \lambda_{11} e^{-\lambda_{11} x_1} \left[\int_0^{x_1} \lambda_{12} e^{-\lambda_{12} x_2} dx_2 \right] dx_1 = \\
&= \int_0^\infty \lambda_{11} e^{-\lambda_{11} x_1} \left[1 - e^{-\lambda_{12} x_1} \right] dx_1 = \\
&= \int_0^\infty \lambda_{11} e^{-\lambda_{11} x_1} \left[1 - e^{-\lambda_{12} x_1} \right] dx_1 = \\
&= 1 - \int_0^\infty \lambda_{11} e^{-(\lambda_{11} + \lambda_{12}) x_1} dx_1 = 1 - \frac{\lambda_{11}}{\lambda_{11} + \lambda_{12}} = \frac{2}{3}
\end{aligned}$$

2 Determine for each of the following statements whether it is *true* or *false*.
No motivation is required, but you should be certain about your choice. For each statement, a correct answer gives +1 point, no answer gives 0 points, but an incorrect answer gives −1 point! A negative sum in this problem will count as 0. The final result can be any integer from 0 to the maximum of (5p).

(a) A probability density function for a K-dimensional feature vector $\boldsymbol{X}$, of the Gaussian mixture model (GMM) type with M Gaussian components,

$$f_{\boldsymbol{X}}(\boldsymbol{x}) = \sum_{m=1}^{M} w_m \frac{1}{(2\pi)^{K/2} \sqrt{\det C_m}} e^{-\frac{1}{2}(x - \boldsymbol{\mu}_m)^T C_m^{-1} (x - \boldsymbol{\mu}_m)}$$

with *diagonal* covariance matrices C_m, can only model feature vectors $\boldsymbol{X}$ with statistically *independent* elements X_k.
Solution: FALSE. If the mean vectors $\boldsymbol{\mu}_m$ are different for different m, the GMM can model many dependencies between feature elements. For example, in $K = 2$ dimensions, we can place the GMM mean vectors along the line $x_1 = x_2$, and then X_1 and X_2 are clearly correlated.

(b) In a *left-right* hidden Markov model (HMM), the *state duration* D_n (i.e., the number of consecutive time instances t where the state remains at $S_t = n$), is a random variable that approaches a Gaussian distribution, if the total number of states, N, is very large, and $1 << n << N$.
Solution: FALSE. In a regular HMM the state duration D_n always has a geometric distribution, with the maximum probability for $D = 1$.

(c) You have previously trained a Gaussian density function on a training set of scalar feature values. Now you need to modify the feature extractor so that all numerical feature values in the training set are scaled by a factor $c > 1$. If you then re-train the Gaussian density function with the scaled training data, all probability-density values will be increased by the same factor c.
Solution: FALSE. If features are transformed as $Y = cX$, the probability density $f_Y(y) = \frac{1}{c} f_X(x)$, because the integral of the density equals 1 in both cases.

(d) A Markov chain with the following initial state probabilities and state transition probabilities is *ergodic*:

$$\text{Initial prob.:} \quad q = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}; \quad \text{Transition prob.:} \quad A = \begin{pmatrix} 0.9 & 0.1 & 0 \\ 0.5 & 0 & 0.5 \\ 0.1 & 0 & 0.9 \end{pmatrix};$$

Solution: TRUE. The Markov chain is ergodic because it is irreducible and aperiodic.

(e) Given an observed output sequence $\underline{x} = (x_1, \ldots, x_T)$ from a hidden Markov model λ, we can use the results of the *Viterbi* algorithm to calculate the conditional state probability

$$P\left(S_t = i | (x_1, \ldots, x_t), \lambda\right)$$

for any i and any $1 \leq t \leq T$.
Solution: FALSE. The Viterbi algorithm can only find the most probable state sequence, and the probability of that particular sequence.

3 You can observe some elements of the output sequence $\boldsymbol{x} = (x_1, \ldots, x_t, \ldots)$ from a discrete hidden Markov source, but you do not know the corresponding internal state sequence $\boldsymbol{S} = (S_1, \ldots, S_t, \ldots)$ in the source. The initial state probability vector is

$$q = \begin{pmatrix} 0.2 \\ 0.8 \end{pmatrix}, \text{with elements } P(S_1 = i).$$

The state transition probability matrix is

$$A = \begin{pmatrix} 0.6 & 0.4 \\ 0.1 & 0.9 \end{pmatrix}, \text{with elements } a_{ij} = P(S_{t+1} = j | S_t = i).$$

The output probability matrix is

$$B = \begin{pmatrix} 0.1 & 0.4 & 0.5 \\ 0.7 & 0.2 & 0.1 \end{pmatrix}, \text{with elements } b_{ik} = P(X_t = k | S_t = i).$$

(a) Calculate $P(X_2 = 3\ )$. (2p)

Solution: The Markov chain is stationary; the stationarity is verified by showing that $A^T q = q$.

Therefore, the unconditional state probabilities are $P(S_t = i) = P(S_1 = i) = q_i$, for any t.

$$P(X_2 = 3) = \sum_i P(X_2 = 3 \cap S_2 = i) =$$
$$= P(X_2 = 3|S_2 = 1)P(S_2 = 1) + P(X_2 = 3|S_2 = 2)P(S_2 = 2) =$$
$$= b_{13}q_1 + b_{23}q_2 = 0.5 \cdot 0.2 + 0.1 \cdot 0.8 = 0.18$$

(b) Calculate $P(S_3 = 1|S_1 = 1 \cap X_1 = 3 \cap X_2 = 3 \cap S_4 = 1 \cap X_4 = 3)$. (3p)

Solution: Given S_1 and S_4, S_3 is statistically independent of X_1 and X_4, so we have

$$P(S_3 = 1|S_1 = 1 \cap X_1 = 3 \cap X_2 = 3 \cap S_4 = 1 \cap X_4 = 3) =$$
$$= P(S_3 = 1|S_1 = 1 \cap X_2 = 3 \cap S_4 = 1) =$$
$$= \frac{P(X_2 = 3 \cap S_3 = 1 \cap S_4 = 1|S_1 = 1)}{P(X_2 = 3 \cap S_3 = 1 \cap S_4 = 1|S_1 = 1) + P(X_2 = 3 \cap S_3 = 2 \cap S_4 = 1|S_1 = 1)}$$

To calculate the probabilities needed in this expression, we first note the two possibilities for S_2, and calculate

$$P(S_2 = i \cap X_2 = 3 \cap S_3 = 1 \cap S_4 = 1|S_1 = 1) = a_{1i}b_{i3}a_{i1}a_{11}$$
$$P(S_2 = i \cap X_2 = 3 \cap S_3 = 2 \cap S_4 = 1|S_1 = 1) = a_{1i}b_{i3}a_{i2}a_{21}$$

Summing over the two possible values for S_2, we obtain the two desired probabilities as

$$P(X_2 = 3 \cap S_3 = 1 \cap S_4 = 1|S_1 = 1) = a_{11}b_{13}a_{11}a_{11} + a_{12}b_{23}a_{21}a_{11}$$
$$P(X_2 = 3 \cap S_3 = 2 \cap S_4 = 1|S_1 = 1) = a_{11}b_{13}a_{12}a_{21} + a_{12}b_{23}a_{22}a_{21}$$

Thus,

$$P(S_3 = 1|S_1 = 1 \cap X_2 = 3 \cap S_4 = 1) =$$
$$= \frac{a_{11}b_{13}a_{11}a_{11} + a_{12}b_{23}a_{21}a_{11}}{a_{11}b_{13}a_{11}a_{11} + a_{12}b_{23}a_{21}a_{11} + a_{11}b_{13}a_{12}a_{21} + a_{12}b_{23}a_{22}a_{21}}$$

4 You have good reasons to assume that the lifetime of your company's products has an exponential distribution, i.e., the lifetime X for any produced item has a conditional probability density function of the form

$$f_{X|W}(x|w) = we^{-wx}, \quad \text{with } E[X|W=w] = \frac{1}{w}, \quad \text{var}[X|W=w] = \frac{1}{w^2}$$

given that the parameter $w > 0$, the failure rate, is exactly known. However, as this parameter is not known, we now regard it as an outcome of a random variable W. In a sample of N items taken at random from the production, you have observed the lifetimes $\underline{x} = (x_1, \ldots, x_N)$. Now you will apply Bayesian estimation to determine the predictive density function for the lifetime X_{N+1} of any future product item, given the observed sequence $\underline{x}$.

(a) Show that the gamma density function is a suitable *conjugate density* form for the parameter W. The gamma density function can be written as

$$f_W(w) = \frac{b^a}{\Gamma(a)} w^{a-1} e^{-bw}, \quad \text{and } E[W] = \frac{a}{b}, \quad \text{var}[W] = \frac{a}{b^2}$$

with hyperparameters $a > 0$ and $b > 0$. (1p)

Solution: Assuming the prior parameter density has the gamma form with hyperparameters a_0, b_0, then the posterior parameter density, given an observed sequence $\underline{x}$, has the form

$$f_{W|\underline{X}}(w|(x_1,\ldots,x_N)) \propto f_{\underline{X}|W}((x_1,\ldots,x_N)|w) f_W(w) = \left[\prod_{n=1}^{N} f_{X|W}(x_n|w)\right] f_W(w) \propto$$

$$\propto \left[\prod_{n=1}^{N} we^{-wx_n}\right] w^{a_0-1} e^{-b_0 w} = w^{a_0+N-1} e^{-w(b_0+\sum_{n=1}^{N} x_n)}$$

(Here, we have also assumed that the different observations x_n are conditionally independent, given w.) Thus, the posterior density has again the gamma form, which is the requirement for a *conjugate density*. The posterior hyperparameters are

$$a_N = a_0 + N; \quad b_N = b_0 + \sum_{n=1}^{N} x_n$$

(b) Determine the non-informative Jeffreys prior density function $f_W(w)$ for the parameter W and express the result in terms of hyperparameters a_0 and b_0 for the gamma density function. (1p)
Hint: Jeffreys prior is defined as

$$f_W(w) \propto \sqrt{E_X\left[\left(\frac{\partial \ln f_{X|W}(X|w)}{\partial w}\right)^2\right]}$$

Solution: Following the definition of Jeffreys prior, we have

$$\begin{aligned}
\frac{\partial \ln f_{X|W}(X|w)}{\partial w} &= \frac{\partial(\ln w - wX)}{\partial w} = \frac{1}{w} - X = E_X\left[X|w\right] - X \\
\left(\frac{\partial \ln f_{X|W}(X|w)}{\partial w}\right)^2 &= \left(\frac{1}{w} - X\right)^2 = \left(E_X\left[X|w\right] - X\right)^2 \\
E_X\left[\left(\frac{\partial \ln f_{X|W}(X|w)}{\partial w}\right)^2\right] &= E_X\left[\left(E_X\left[X|w\right] - X\right)^2 |w\right] = \mathrm{var}\left[X|w\right] = \frac{1}{w^2}
\end{aligned}$$

Thus, Jeffreys prior has the form

$$f_W(w) \propto \frac{1}{w}$$

This can be seen as a gamma density with hyperparameters $a_0 \to 0;\ b_0 \to 0$.

(c) Determine the predictive density function $f_{X_{N+1}|\underline{X}}\left(x|\left(x_1, \ldots, x_N\right)\right)$. (3p)
Hint: If you could not determine the non-informative Jeffreys prior density, it is allowed to use any values for the prior gamma hyperparameters a_0 and b_0 here.

Solution: We have already shown in (a), that the posterior parameter density is a gamma density with hyperparameters

$$a_N = a_0 + N; \quad b_N = b_0 + \sum_{n=1}^{N} x_n$$

Including the normalization factors, we have

$$f_{W|\underline{X}}(w|\underline{x}) = \frac{b_N^{a_N}}{\Gamma(a_N)} w^{a_N - 1} e^{-b_N w}$$

The predictive density for any single future observation is then

$$\begin{aligned}
f_{X_{N+1}|\underline{X}}(x|\underline{x}) = \int_0^\infty f_{X|W}(x|w) f_{W|\underline{X}}(w|\underline{x}) = \\
= \int_0^\infty w e^{-wx} \frac{b_N^{a_N}}{\Gamma(a_N)} w^{a_N - 1} e^{-b_N w} dw = \\
= \frac{b_N^{a_N}}{\Gamma(a_N)} \int_0^\infty w^{a_N + 1 - 1} e^{-(b_N + x)w} dw = \frac{b_N^{a_N}}{\Gamma(a_N)} \frac{\Gamma(a_N + 1)}{(b_N + x)^{a_N + 1}}
\end{aligned}$$

Here, the last integral is found simply by observing that it must be the inverse of the normalization constant for a gamma density with hyperparameters $a_N + 1$ and $b_N + x$. The result can be further simplified by using the recursive relation $\Gamma(z+1) = z\Gamma(z)$ for the gamma function:

$$f_{X_{N+1}|\underline{X}}(x|\underline{x}) = \frac{a_N}{b_N}\left(\frac{b_N}{b_N + x}\right)^{a_N + 1} = \frac{a_N}{b_N}\left(1 + \frac{x}{b_N}\right)^{-a_N - 1}$$

Using the non-informative prior hyperparameters $a_0 \to 0$ and $b_0 \to 0$, it may be instructive to express the result in terms of the observed average lifetime

$$\bar{x} = \frac{\sum_{n=1}^{N} x_n}{N} = \frac{b_N}{a_N}$$

Then the predictive distribution can also be written as

$$f_{X_{N+1}|\underline{X}}(x|\underline{x}) = \frac{1}{\bar{x}} \left(1 + \frac{x}{N\bar{x}}\right)^{-N-1}$$

For large N, this function approaches again the exponential form, with the parameter $w \to 1/\bar{x}$:

$$\lim_{N\to\infty} f_{X_{N+1}|\underline{X}}(x|\underline{x}) = \frac{1}{\bar{x}} e^{-x/\bar{x}}$$

5 A sequence of scalar random values $(X_1, \ldots, X_t, \ldots)$ is generated by the algorithm

$$X_t = cZ_tW_t$$

Here, c is a real-valued constant. The Z_t and W_t values cannot be observed directly. W_t is for every t a Gaussian random variable with mean 0 and variance 1. The random sequence $\underline{Z} = (Z_1, \ldots, Z_t, \ldots)$ contains discrete elements Z_t that can be either $Z_t = 1$ or $Z_t = 2$. All W elements are statistically independent of all Z elements. All W_t values are statistically independent across different t, but the Z_t values have the following conditional probability mass distribution:

$$\begin{aligned} P(Z_1 = 1) =& 1 \\ P(Z_{t+1} = 2|Z_t = 1) =& 2r \\ P(Z_{t+1} = 1|Z_t = 2) =& r \end{aligned}$$

The constant r is exactly known, but c is initially known only as a crude approximation. You have observed a sequence $\underline{x} = (x_1, \ldots, x_t, \ldots, x_T)$ generated by this source. As the source can be described as an HMM, we assume you have already used the forward-backward algorithms to determine the conditional state probabilities at any t, given the observation and the previous parameter value c:

$$\begin{aligned} \gamma_{1,t} =& P(Z_t = 1|\underline{x}, c) \\ \gamma_{2,t} =& P(Z_t = 2|\underline{x}, c) = 1 - \gamma_{1,t} \end{aligned}$$

Now regard all the $\gamma_{i,t}$ values as known, and apply the EM algorithm to determine an update formula to obtain an improved estimate c^{new}, given the initial approximate value c and the observed sequence $\underline{x} = (x_1, \ldots, x_T)$. (5p)
Hint: Each step in the EM algorithm should maximize the help function

$$Q(c', c) = E_{\underline{Z}} \left[\ln P(\underline{Z}, \underline{x}|c')|\underline{x}, c \right]$$

Solution: The EM help function is

$$Q(c', c) = E_{\underline{Z}} \left[\ln P(\underline{Z}, \underline{x}|c')|\underline{x}, c \right] =$$
$$= \sum_{z_1=1}^{2} \cdots \sum_{z_T=1}^{2} P(\underline{Z} = (z_1, \ldots, z_T)|\underline{x}, c) \cdot \ln P(\underline{Z} = (z_1, \ldots, z_T) \cap \underline{x}|c')$$

The $\underline{Z}$ sequence results from a Markov chain with known initial probabilities q_i and transition probabilities a_{ij}, given in terms of r. Thus, we can express the probability for any specific $\underline{Z}$ and $\underline{x}$ sequences as

$$P(\underline{Z} = (z_1, \ldots, z_T) \cap \underline{x}|c') = q_{z_1} b_{z_1}(x_1) a_{z_1 z_2} b_{z_2}(x_2) \cdots a_{z_{T-1} z_T} b_{z_T}(x_T)$$

Here only the b_{z_t} factors depend on the unknown parameter c'. The state-conditional density functions for X_t are Gaussian, with zero mean, and

$$\text{var}\left[X_t|Z_t = 1, c'\right] = c'^2; \quad \text{var}\left[X_t|Z_t = 2, c'\right] = 4c'^2$$

Thus, the density functions are

$$b_1(x) = \frac{1}{\sqrt{2\pi}c'} e^{-x^2/2c'^2}; \qquad b_2(x) = \frac{1}{\sqrt{2\pi}2c'} e^{-x^2/8c'^2}$$

We note that all the factors that do not depend on c' only contribute a constant term in $Q(c', c)$. Therefore, maximizing Q is the same as maximizing

$$\begin{aligned}
q(c', c) = \sum_{z_1=1}^{2} \cdots \sum_{z_T=1}^{2} P(\underline{Z} = (z_1, \ldots, z_T) | \underline{x}, c) \cdot \sum_{t=1}^{T} \ln b_{z_t}(x_t) = \\
= \sum_{t=1}^{T} \sum_{z_t=1}^{2} \underbrace{P(Z_t = z_t | \underline{x}, c)}_{=\gamma_{z_t,t}} \ln b_{z_t}(x_t) \cdot \\
\cdot \underbrace{\sum_{z_1=1}^{2} \cdots \sum_{z_{t-1}=1}^{2} \sum_{z_{t+1}=1}^{2} \cdots \sum_{z_T=1}^{2} P((z_1, \ldots, z_{t-1}, z_{t+1}, \ldots, z_T | Z_t = z_t, c)}_{=1} = \\
= \sum_{t=1}^{T} \gamma_{1,t} \ln b_1(x_t | c') + \gamma_{2,t} \ln b_2(x_t | c') = \\
= \sum_{t=1}^{T} \gamma_{1,t} \left(-\ln c' - x^2/2c'^2 \right) + (1 - \gamma_{1,t}) \left(-\ln c' - x^2/8c'^2 \right) + \text{const.} = \\
= -T \ln c' - \frac{1}{2c'^2} \sum_{t=1}^{T} \gamma_{1,t} x_t^2 + \frac{1 - \gamma_{1,t}}{4} x_t^2 + \text{const.}
\end{aligned}$$

A necessary condition for maximum is

$$0 = \frac{\partial q(c', c)}{\partial c'} = -\frac{T}{c'} + \frac{1}{c'^3} \sum_{t=1}^{T} \left(\frac{1}{4} + \frac{3\gamma_{1,t}}{4} \right) x_t^2$$

This equation has the solution

$$c'^2 = \frac{1}{T} \sum_{t=1}^{T} \left(\frac{1}{4} + \frac{3\gamma_{1,t}}{4} \right) x_t^2$$

which is the desired update equation.

This result makes sense, intuitively. Whenever $\gamma_{1,t}$ is near 1, the observed x_t^2 contributes with weight 1 to the variance which is c'^2 if $Z_t = 1$. Whenever $\gamma_{1,t}$ is near 0, i.e., $\gamma_{2,t}$ is near 1, the observed x_t^2 contributes with weight 1/4 to c'^2, and, thus, with weight 1 to the variance which is $4c'^2$ when $Z_t = 2$.